UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Un Modelo de Regresión Poisson no Lineal para Extinción de Especies en Hábitats Fragmentados

Qitian Lai

Universidad Santo Tomás
Facultad de Estadística
División de Ciencias Económicas y Administrativas
Bogotá, D.C., Colombia
2022

A Nonlinear Poisson Regression Model For Species Extinction in Fragmented Habitats

Qitian Lai

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magister en Estadística Aplicada

Director(a):
Ph.D. Mario José Pacheco López

Línea de Investigación:
Biostatistics
Grupo de Investigación:
USTAdistica

Universidad Santo Tomás
Facultad de Estadística
División de Ciencias Económicas y Administrativas
Bogotá, D.C., Colombia
2022

Acknowledgement

I am very grateful to my supervisor, Mario José, for all his support. He guided me from scratch to learn the skills of writing papers and fill in the knowledge that I lacked as a graduate student in statistics.
I am very grateful to the instructor Andrés Cruz for all the support from my admission to the present. He and all the teachers have taught me from a beginner in statistics to the current level.
I am very grateful to all the teachers and partners who taught me for their support in my course, thesis and life.
Thanks a lot to everyone.

Resumen

La bioestadística es el conjunto de métodos estadísticos utilizados principalmente en biología, incluido el diseño de experimentos biológicos, la recopilación de datos experimentales y el análisis estadístico de estos tipos datos. Un gran número de biólogos recopilan diferentes tipos de datos, de diferentes rincones del mundo. Generalmente, estos datos contienen diferentes tipos de variables dependientes e independientes que pueden analizarse utilizando modelos de regresión lineales o no lineales. En la actualidad, las actividades humanas (construcción de presas, deforestación, etc.) y el cambio climático representan una amenaza para un gran número de especies animales y plantas, y la extinción de las especies representa una gran parte de la investigación biológica y bioestadística. En cuanto a la modelización estadística de la extinción de especies y sus posibles causas, es común encontrar en propuesta de Regresión Normal No Lineal, Lo cual reducir errór, dado que la variable respuesta corresponde al número de especies en ciertas áreas de interés a lo largo del tiempo, por lo que la Regresión de Poisson no Lineal sería una suposición más apropiada. En este proyecto, se propone la Regresión de Poisson No Lineal con el objetivo de ajustar los datos sobre el número de especies en hábitats fragmentados. En particular, tiene como objetivo refinar el Modelo de Gibson et al. (2013), que es un modelo No Lineal para la variable número de especies, en función del área de los fragmentos y la fragmentación del hábitat. Este proyecto demuestra que la Regresión de Poisson No Lineal es más eficiente que otros tipos de Regresiones, al ajustar tipos de datos comunes en bioestadísticas como el número de especies en función del tiempo y el área de los fragmentos. Los resultados mostraron que existe una correlación negativa entre el número de especies restantes y el tiempo. A medida que pasa el tiempo, el número de especies restantes disminuye, existe una correlación positiva entre el número de especies restantes y el área de la isla. Las islas más grandes tienen más especies restantes, pero con el paso del tiempo, la tasa de extinción de las islas más grandes es más alta que la de las islas pequeñas. Sin embargo al final las especies restantes en la isla eventualmente se extinguirán hasta la situación en la que solo quedarán una o dos especies que básicamente tienden a extinguirse hasta 15 año. Para usar Regresión Normal no Lineal y Regresión de Poisson no Lineal respectivamente, comparando la Regresión Original de luke et al. (2013), el valor R^2 también es más alto que las otras dos regresiones, por otro lado, la Regresión de Poisson no Lineal tiene un intervalo de confianza más pequeño y menor el desviación residual y AIC. Se puede concluir que la Regresión Poisson no lineal es más eficiente que la Regresión Normal no Lineal.

Palabras clave: Bioestadística; Extinción de Especies; Fragmentación de Hábitats; Modelo de Regresión Normal no Lineal; Modelo de Regresión Poisson no Lineal.

Abstract

Biostatistics is the set of statistical methods used mainly in biology, including the design of biological experiments, the collection of experimental data, and the statistical analysis of these data. Numerous biologists collect different data, from different corners of the world. These data contain different dependent and independent variables which can be analyzed using linear, or nonlinear regression models. At present, the human activities (dam building, deforestation, etc.) and climate change represent a threat to numerous animal and plant species, and the extinction of the species represents a large part of biological and biostatistical research. Regarding the statistical modeling of the extinction of species and its causes, it is common to find out most of the research are using the Nonlinear Normal Regression, this kind of Regression is mistakenly assumed, the response variable corresponds to the number of species in certain areas of interest with time passing, thus Nonlinear Poisson regression would be a more appropriate assumption. In this project, the Nonlinear Poisson regression is proposed, aiming to fit data concerning the number of species in fragmented habitats. In particular, it aims to refine the Island Biogeographic Model in Gibson et al. (2013), which is a Nonlinear model for the variable number of species, as a function of the area of the fragments and habitat fragmentation. This project proves that Nonlinear Poisson Regression is more effective than other types of regressions. When fitting the data in biostatistics, the variable includes the number of species, time, and the area of fragments. The results showed that there have a negative correlation between the number of remaining species and time, with time passing, the number of remaining species is decreasing, there is a positive correlation between the number of remaining species and the island area, larger islands have more number of remaining species, but with the time passing, the extinction rate of larger islands is higher than small islands, however by the end, the remaining species on the island will eventually become extinct to the situation where only have one or two species remaining. For using Nonlinear Normal Regression and Nonlinear Poisson Regression respectively, comparing the regression original from Luke et al. (2013), the R^2 value is also higher than the other two regressions. In other hand, the Nonlinear Poisson Regression has a smaller confidence interval and residual deviance and AIC. We can concluded that when using Nonlinear Poisson Regression, the regression analysis of data is more effective than using the Nonlinear Normal Regression.

Keywords: Biostatistics; Habitat Fragmentation; Nonlinear Normal Regression Model; Nonlinear Poisson Regression Model; Species Extinction.

List of Figures

List of Tables

1 Introduction

Biostatistics is the branch of statistics mainly applied to the collection, analysis and interpretation of biological data, at present human activities (dam building, deforestation, etc.) and climate change, represent a threat to numerous animal and plant species, and the extinction of the species makes up a large part of the biological and biostatistical research. For example, Megan et al. (2010) conducted that a study to determine the relative importance of mosaic landscape properties for native terrestrial mammal species richness, in highly changed landscapes. Aguilar et al. (2018) studied that losing plant species after a decade in fragmented subtropical forests. Palmeirim et al. (2018) evaluated that small mammal assemblage responses to Amazonian forest habitat insularization induced by the construction of a dam after 28 years. Newmark et al. (2017) assessed that the impact of targeted habitat restoration, and estimated the persistent increase in the number of species by regenerating a forest connection.

In many biostatistical papers, it was shown that with environmental destruction, the population of species shows a significant decline, in some time even extinct to zero, for example Michalski et al. (2007) showed that human logging activities in the Amazon rain-forest leads to the division of a species into areas, then rapidly become extinct with time passing. In fact, something closely related the number of species extinctions as time and area, and geographic location. In particular, Gibson et al. (2013) studied that near-complete extinction of native small mammal fauna 25 years after forest fragmentation, and proposed a Nonlinear Normal Regression model called Island Biogeographic Model as below.

$$S(t) = s_\infty - (s_\infty - ca^z)e^{-kt}$$

Where $S(t)$ is the number of species at time t, a is the area of the fragment (island), and t is the time since isolation. s_∞, c, k and z are the parameters of the model.

A problem in the Gibson et al. (2013) model is that the response variable was assumed to be absolutely continuous, and following Normal Regression instead of Poisson Regression. This means that the entire fitting procedure was based on a Normal Nonlinear Regression Model. In this study, a Poisson Nonlinear Regression model is proposed for the number of species at a specific time t, $S(t)$ in The Island Biogeographic Model, allowing a better estimation of the model parameters, so a better fitting result is to be expected. The following approach will be used. First we will use the Nonlinear Normal Regression and Nonlinear Poisson Regression separately to estimate the number of species at time t, then we will compare the results got

from both regressions, to verify that the Nonlinear Poisson Regression can process the data more effectively.
In order to facilitate the calculation, the original formula needs to be redefined.
From

$$S(t) = s_\infty - (s_\infty - ca^z)e^{-kt}$$

to

$$S(t) = s_\infty(1 + (we^{zIna} - 1)e^{-kt}))$$

After using the Nonlinear Normal Regression redefined, the result shows a higher coefficient of determination (R^2) than the result given in the original article Gibson et al. (2013). So when using the Nonlinear Poisson Regression for fitting, the result shows it is more effective than using the Nonlinear Normal Regression for fitting, even if the value of coefficient of determination is almost the same, it has a lesser AIC value and the residual degrees of freedom.
The above comparison leads to the conclusion is that for this type of biological experiment, the using of Nonlinear Poisson Regression can more effectively estimate the data model of the number of species in the area, time and area of the fragment.
We can extend the conclusion that for this type of biological data, especially data with variables such as island area (we can also call it isolation area) and time, the Nonlinear Poisson Regression can be more effective than the other regression. Consequently, the used fitting regression for this type of data can be discussed.

2 Theoretical Framework and Literature Review

2.1. Biodiversity

From UNEP-WCMC (2010) we can find the definition of Biodiversity, in biology, the diversity of species is an important research field. With economic development, human beings are increasingly making a stronger impact on nature. Because of this influence, there are environmental problems made by some economic activities, responsible for serious threats to species diversity. This impact leads to the extinction of species, which is currently the most important problem addressed by environmental protection activities. Biodiversity is the integration of ecological complexes formed by organisms, their environments and multiple ecological processes related to them. It includes animals and plants and microorganisms and their own genes, as well as complex ecosystems formed by their living environments.
Species diversity usually includes three components are genetic diversity, species diversity, and ecosystem diversity. Genetic diversity can ensure species diversity, species diversity can ensure ecological diversity. Species diversity is the abundance of biological species such as animals, plants, and microorganisms on the earth.
When mention about Species diversity, this includes two parts, one refers to the richness of species in a certain area, which can be called regional species diversity, the other refers to the uniformity of species distribution in ecology, also called ecological diversity or community Species diversity. Species diversity is an aim indicator to measure the abundance of biological resources in a certain area. Therefore, when describing the abundance of biodiversity in a region, the most commonly used indicator is regional species diversity.
The measurement of regional species diversity has the following three indicators which are the first one is the total number of species, this is the number of species of a specific group, in a specific area, the second one is species density which refers to the number of species of a specific group, per unit area, the third one is proportion of endemic species which refers to the proportion of endemic species of a particular group in a certain area, related to the total number of species in the area. Additionally, numerous articles are also leading this investigation, and about 40 % of biology articles published in the past 20 years are part of this investigation.
Most articles showed that biodiversity is affected by time, area size, region, and ecological community. For example, Megan et al. (2010) showed that in isolated ecological areas, the

diversity of mammals and plants is based on the size of the area and the environment in which they are located. Aguilar et al. (2018) showed that in the independent ecological communities formed after the disintegration of forests, it positively correlated the number of residual species to the area of the ecological community. Palmeirim et al. (2018) showed that the destruction of the Amazon rain-forest because hydro-power dams, in which the 0.667 of the new islands formed, were likely to be a subject to biological extinction. The smaller island is, the more likely is to lead to species extinction. Newmarka et al. (2017) showed that in a broken ecosystem, it takes an average of 6.7 years to restore species, and restoring species diversity requires spending a large amount of money. Benchimol et al. (2014) and Sharma et al. (2014) showed that large mammals are most likely to become extinct in a fragmented ecosystem. Sodhi et al. (2013) showed that with the increase in human activities, including logging and poaching, the species of tropical islands have been dwindling.
Mendes et al. (2016) showed that a generalized Nonlinear Normal Regression Model designed to analyze to what extent the vegetation type, fragment area, isolation, sampling effort (as total kilometers), the correlation of each variable, and then with the date it can predict the number of richness and extinction rates. Megan et al. (2011) showed that in Australia there have a large amount of species in the designed area, so that in the designed area there is a greater population of mammals. Martensen et al. (2012) and Aguilar et al. (2018) showed that in Southeastern Brazil, as the forest disappears, the species have more abilities to adapt to the environment, can multiply their population rapidly, while the number of species not able to adapt to the environment quickly declines. Chisholm et al. (2018) showed that the use of new estimation methods about Nonlinear Poisson Regression to estimate the status of species extinction.
There are many reasons for the extinction of species, one is the impact of the isolation of the island, and the other is the invasion of rodents such as Malayan field rat (Rattus tiomanicus) which is a species of rodent in the family Muridae. It is nocturnal, mainly arboreal, and is found in Malaysia, Thailand, Indonesia and Philippines. However, Luke et al. (2013) is an environmental expert not a statistician, so his calculations are problematic. This is the reason we need to analyze the regression of the date to find a better way to estimate the model.

2.2. Island Biogeographic Model

The authors of Gibson et al. (2013) derived an island biogeographic model to predict the number of species on forest fragments that emerged after island isolation in order to better perform regression analysis of the data. The equilibrium number of species on an island is assumed to follow a power-law model

$$S_0 = ca^z$$

For the parameter S_0 is the initial number of species on the island before isolation, a is the area of each island, and c and z are constants of number. Simple power-law species-area

relationships could perform best across the data
The theory of island biogeography could show that the change in the number of species on an island would be

$$S_{t+1} = S_t + I - E$$

where S_{t+1} and S_t are the number of species at times $t+1$ and t, in another hand I is the number of new species immigrating to an island during the elapsed time interval (t, $t+1$), and E is the number of extinctions (including permanent emigration) on an island during the elapsed time interval. If define the parameters I and E. We could get some examples, they can be functions of island size and the number of resident species on the island. The number of parameters can quickly increase if we consider both area and number of species for each parameter. For this time, we consider one simple model as below:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = I_0(S_m - S) - E_0 S$$

For the parameter S_m is the species pool on the mainland, and E_0 and I_0 are extinction and immigration rates. We could get as below:

$$S_t = \frac{I_0 S_m}{I_0 + E_0} + (\frac{I_0 S_m}{I_0 + E_0} - S_0)e^{-(I_0+E_0)t}$$

where S_0 is the number of richness on an island before isolation, for the model $S_0 = ca^z$ is the equilibrium number of species of the original system. Substituting the model $S_0 = ca^z$ into the equation above and simplifying notation. Finally The Island Biogeographic Model proposed by Gibson et al. (2013) is defined by

$$S(t) = s_\infty - (s_\infty - ca^z)e^{-kt}$$

Where $S(t)$ is the number of species at time t, a is the area of the fragment (or island), and t is time since isolation (fragmentation), so that s_∞, c, k and z are the parameters of the model.
A fundamental aspect in this model is the interpretation of the parameter s_∞, which corresponds to the number of species in relaxation (infinity, when $t \to \infty$), understanding relaxation as the gradual loss of species through time after a habitat is fragmented.
Note that we can see The Island Biogeographic Model as a particular case of the General Nonlinear Model

$$Y_i = f(\mathbf{x}_i, \theta) + \boldsymbol{\varepsilon}_i$$

Where Y_i is the response variable, $\mathbf{x}_i$ are the explanatory variables, θ are the parameters, and $\boldsymbol{\varepsilon}_i$ is the error term, defining Y_i as $S(t)$, and $f(\mathbf{x}_i, \theta)$ as $s_\infty - (s_\infty - ca^z)e^{-kt}$ we can get the formula as below:

$$S(t) = s_\infty - (s_\infty - ca^z)e^{-kt}$$

Formula Derivation

From the Island Biogeography Model proposed by Gibson et al. (2013), the original formula can be got as it follows:

$$S(t) = s_\infty - (s_\infty - ca^z)e^{-kt}$$

As the same variable is included in the formula, it is difficult to calculate the data. Therefore, we need to be redefined the formula to make it easier to calculate the data.
We start the conversion of the formula

$$\begin{aligned} S(t) &= s_\infty - (s_\infty - ca^z)e^{-kt} \\ &= s_\infty + (ca^z - s_\infty)e^{-kt} \\ &= s_\infty + (ce^{zIna} - s_\infty)e^{-kt}. \end{aligned}$$

We define $s_\infty = c/w$, so we can redefine the formula as it follows

$$\begin{aligned} S(t) &= s_\infty - (s_\infty - ca^z)e^{-kt} \\ &= s_\infty(1 - (s_\infty - ca^z)e^{-kt}/s_\infty)) \\ &= s_\infty(1 - (1 - ca^z/s_\infty)e^{-kt})) \\ &= s_\infty(1 - (1 - wa^z)e^{-kt})) \\ &= s_\infty(1 + (wa^z - 1)e^{-kt})) \\ &= s_\infty(1 + (we^{zIna} - 1)e^{-kt})) \end{aligned}$$

Where $w = c/s_\infty$. The final formula is got, and makes it easier to use the R program language to estimate the data. Then we need to define the concept of Nonlinear Normal Regression and Nonlinear Poisson Regression.

2.3. Concept based on Nonlinear Normal Regression

In Multivariate nonlinear regression model, we can get the following relation:

$$Y_i \sim N(f(x_i, \theta), \sigma^2)$$

Where Y_i is the vector of response variable, x_i is the vector of explanatory variable, θ is the parameters vectors included s_∞, σ, c, k and z.

$$f(x_i, \theta) = s_\infty(1 + (we^{zIna} - 1)e^{-kt})$$

Where $\theta = (s_\infty, \sigma, w, k, z)$.
We put the formula into Multivariate Nonlinear Regression, for normal distribution $\mathcal{N}(\mu, \sigma^2)$ and has the following function:

$$f(y_i \mid \theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - f(x_i, \theta))^2}{2\sigma^2}\right),$$

For the formula of n from the date (the likelihood) is as it follows:

$$\begin{aligned}
\mathcal{L}(y_1, \ldots, y_n \mid \theta, \sigma^2) &= \prod_{i=1}^{n} f(y_i \mid \theta, \sigma^2) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - f(x_i, \theta))^2}{2\sigma^2}\right) \\
&= \left(\frac{1}{(2\pi)^{n/2}\sigma^n}\right) \exp\left(-\frac{\sum_{i=1}^{n}(y_i - f(x_i, \theta))^2}{2\sigma^2}\right)
\end{aligned}$$

This family of distributions has various parameters $\theta = (s_\infty, \sigma, w, k, z)^\top$, so we maximize the likelihood, $\mathcal{L}(\theta, \sigma) = f(x_1, \ldots, x_n \mid \theta, \sigma)$, over all parameters simultaneously or if possible individually.

Maximum Likelihood Estimation of the Island Biogeographic Model based on the Nonlinear Normal Regression

Maximum likelihood estimation (MLE) is an important and common method for obtaining estimators. The maximum likelihood method explicitly uses a model, and its goal is to find a phylogenetic tree being able to produce observation data with a higher probability.

Since the logarithm function itself is a continuous, strictly increasing function over the range of the likelihood, the values maximizing the likelihood will also maximize its logarithms. The log-likelihood can be written as it follows:

$$\begin{aligned}
log\mathcal{L}(x_1, \ldots, x_n \mid \theta, \sigma) &= \frac{-nlog(2\pi)}{2} - \frac{nlog(\sigma^2)}{2} - \frac{1}{(2\sigma^2)} \sum_{i=0}^{n} (y_i - f(x_i, \theta))^2 \\
&= \frac{-nlog(2\pi)}{2} - \frac{nlog(\sigma^2)}{2} - \frac{1}{(2\sigma^2)} \sum_{i=0}^{n} (y_i - s_\infty(1 + (we^{zIna} - 1)e^{-kt}))^2
\end{aligned}$$

To maximize its logarithms need to maximize for relation of θ, because θ are the vector parameters included s_∞, σ, w, k, z. So that we need to separate to each parameter, to find the relationship for each parameter.

To Maximize for s_∞, the partial derivative of the logarithms likelihood must be equal to

zero, regarding s_∞, we got the formula as below:

$$\begin{aligned} \frac{\partial log\mathcal{L}(x_1,\ldots,x_n \mid \theta,\sigma)}{\partial s_\infty} &= 0 \\ \frac{1}{2\sigma^2}\sum_{i=0}^{n} -(1+(we^{zIna}-1)e^{-kt}) &= 0 \\ \frac{1}{2\sigma^2}\sum_{i=0}^{n} (1+(we^{zIna}-1)e^{-kt}) &= 0 \end{aligned}$$

For this part we could find out that for $(1+(we^{zIna}-1)e^{-kt})$ is zero, here we could not find out the relationship only with formula, so for this part we need to use R language program to find out the value of each parameter.
To Maximize for w, the partial derivative of the logarithms likelihood must be equal to zero, regarding w we got the formula as below:

$$\begin{aligned} \frac{\partial log\mathcal{L}(x_1,\ldots,x_n \mid \theta,\sigma)}{\partial w} &= 0 \\ \frac{1}{2\sigma^2}\sum_{i=0}^{n} -(s_\infty(e^{zIna})e^{-kt}) &= 0 \\ \frac{1}{2\sigma^2}\sum_{i=0}^{n} (s_\infty(e^{zIna})e^{-kt}) &= 0 \end{aligned}$$

For this part we could find out that for $(s_\infty(e^{zIna})e^{-kt})$ is zero, here we could not find out the relationship only with formula, so for this part we need to use R language program to find out the value of each parameter.
To Maximize for σ, the partial derivative of the logarithms likelihood must be equal to zero, regarding σ we got the formula as below:

$$\begin{aligned} \frac{\partial log\mathcal{L}(x_1,\ldots,x_n \mid \theta,\sigma)}{\partial \sigma} &= 0 \\ -\sigma n log'(\sigma^2) + \frac{1}{\sigma^3}\sum_{i=0}^{n}(y_i - s_\infty(1+(we^{zIna}-1)e^{-kt})) &= 0 \\ \sigma^4 n log'(\sigma^2) - \sum_{i=0}^{n}(y_i - s_\infty(1+(we^{zIna}-1)e^{-kt})) &= 0 \end{aligned}$$

For this part we could find out that for $\sigma^4 n log'(\sigma^2) - \sum_{i=0}^{n}(y_i - s_\infty(1+(we^{zIna}-1)e^{-kt}))$ is zero, here we could not find out the relationship only with formula, so for this part we need to use R language program to find out the value of each parameter.
To Maximize for z, the partial derivative of the logarithms likelihood must be equal to zero,

regarding z we got the formula as below:

$$\begin{aligned} \frac{\partial log\mathcal{L}(x_1,\ldots,x_n \mid \theta,\sigma)}{\partial z} &= 0 \\ \frac{1}{2\sigma^2}\sum_{i=0}^{n} -z(s_\infty w(e^{zIna})e^{-kt}) &= 0 \\ \frac{1}{2\sigma^2}\sum_{i=0}^{n} z(s_\infty w(e^{zIna})e^{-kt}) &= 0 \end{aligned}$$

For this part we could find out that for $z(s_\infty w(e^{zIna})e^{-kt})$ is zero, here we could not find out the relationship only with formula, so for this part we need to use R language program to find out the value of each parameter.

To Maximize for k, the partial derivative of the logarithms likelihood must be equal to zero. Regarding k we got the formula as below:

$$\begin{aligned} \frac{\partial log\mathcal{L}(x_1,\ldots,x_n \mid \theta,\sigma)}{\partial k} &= 0 \\ \frac{1}{2\sigma^2}\sum_{i=0}^{n} (s_\infty(we^{zIna}-1)e^{-kt}) &= 0 \end{aligned}$$

For this part we could find out that for $(s_\infty(we^{zIna}-1)e^{-kt})$ is zero, here we could not find out the relationship only with formula, so for this part we need to use R language program to find out the value of each parameter.

Where $\theta = (s_\infty, \sigma, w, k, z)$ has no closed-form solution, we can get the result concerning MLE base on Normal distribution, from the above formula, and the corresponding relationship of each parameter can be got. Here, we need to use the R language program to estimate each parameter. when using the R language program, we need to use the *nls* and *gnm* packages, take the formula from section 2.2 and choose the nonlinear normal regression format, import the data, and draw conclusions based on the maximize its logarithms of each parameter from the data to estimate the value of each parameter.

Confidence Intervals for Regression Coefficients of the Island Biogeographic Model based on the Nonlinear Normal Regression

Allaire et al. (2021) show that the problem to be solved by the confidence interval coefficient is to examine how close the estimated coefficient value is to the true value of the coefficient. The smaller the confidence interval coefficient means it is more effective expression. For nonlinear normal regression fitting, the coefficient estimation value was got from section 2.3, which is the value of parameters vectors $\widehat{\theta}$ included $\widehat{s}_\infty$, $\widehat{\sigma}$, $\widehat{c}$, $\widehat{k}$ and $\widehat{z}$.

For the confidence interval coefficients of Nonlinear Normal Regression, we can get the formula as follows:

$$\mathrm{CI}^{\widehat{\theta}}_{0,95} = \left[\hat{\theta} - t_{\alpha/2}(n-d-1) \times SE(\hat{\theta})\,,\, \hat{\theta} + t_{\alpha/2}(n-d-1) \times SE(\hat{\theta})\right].$$

In generally the number of parameters we defined as k, but in the regression there have the coefficient k, to avoid the difference between number of parameters k and parameter of $\widehat{k}$, we defined the number of parameters as d. α is 5 %, from t table we got the value of t base on $\alpha/2$ is 0.025, from the date we got that n is 64, d is 4, $n-d-1$ is 59, so that $t_{\alpha/2}(n-d-1)$ is 2.00. The Standard Error (SE) of an estimate of a coefficient is the Standard Deviation of the estimate.

So that we can get the coefficient confidence interval of the vectors $\widehat{\theta}$ included $\widehat{s}_\infty$, $\widehat{\sigma}$, $\widehat{c}$, $\widehat{k}$ and $\widehat{z}$. as it is shown here below

$$\begin{aligned}
\mathrm{CI}_{0,95}^{\widehat{s_\infty}} &= \left[\hat{s}_\infty - t_{\alpha/2}(n-d-1)\times SE(\hat{s}_\infty)\,,\ \hat{s}_\infty + t_{\alpha/2}(n-d-1)\times SE(\hat{s}_\infty)\right].\\
\mathrm{CI}_{0,95}^{\widehat{\sigma}} &= \left[\hat{\sigma} - t_{\alpha/2}(n-d-1)\times SE(\hat{\sigma})\,,\ \hat{\sigma} + t_{\alpha/2}(n-d-1)\times SE(\hat{\sigma})\right].\\
\mathrm{CI}_{0,95}^{\widehat{c}} &= \left[\hat{c} - t_{\alpha/2}(n-d-1)\times SE(\hat{c})\,,\ \hat{c} + t_{\alpha/2}(n-d-1)\times SE(\hat{c})\right].\\
\mathrm{CI}_{0,95}^{\widehat{k}} &= \left[\hat{k} - t_{\alpha/2}(n-d-1)\times SE(\hat{k})\,,\ \hat{k} + t_{\alpha/2}(n-d-1)\times SE(\hat{k})\right].\\
\mathrm{CI}_{0,95}^{\widehat{z}} &= \left[\hat{z} - t_{\alpha/2}(n-d-1)\times SE(\hat{z})\,,\ \hat{z} + t_{\alpha/2}(n-d-1)\times SE(\hat{z})\right].
\end{aligned}$$

When use nonlinear normal regression to estimate regression analysis on each column from the data, for each coefficient aspect to each variable, so here it can be said that the defined the Standard Error (SE) of an estimate of each coefficient, according to Draper Norman et al. (1998)(chapter 24) we got the definition for Nonlinear Normal Regression of the standard error of the regression slope coefficient θ_i is estimated as $se(\theta)$= {appropriate diagonal element of $((\widehat{Z}'\widehat{Z})^{-1}SE_{yx}^2)^{1/2}$ }

Where ${Z_{iu}}^{\widehat{\theta}}$ is to maximize its logarithms need to maximize for relation of θ as below

$${Z_{iu}}^{\widehat{\theta}} = \left[\frac{\partial log\mathcal{L}(x_1,\ldots,x_n \mid \theta,\sigma)}{\partial\theta}\right]_\theta$$

Where $\widehat{Z}$ is the Z of estimated value

$$Z_{\widehat{\theta}} = \begin{bmatrix}
{Z_{11}}^{\widehat{\theta}} & {Z_{21}}^{\widehat{\theta}} & . & . & . & {Z_{p1}}^{\widehat{\theta}}\\
{Z_{12}}^{\widehat{\theta}} & {Z_{22}}^{\widehat{\theta}} & . & . & . & {Z_{p2}}^{\widehat{\theta}}\\
. & . & & & & .\\
. & . & & & & .\\
. & . & & & & .\\
{Z_{1u}}^{\widehat{\theta}} & {Z_{2u}}^{\widehat{\theta}} & . & . & . & {Z_{pu}}^{\widehat{\theta}}\\
. & . & & & & .\\
. & . & & & & .\\
. & . & & & & .\\
{Z_{1n}}^{\widehat{\theta}} & {Z_{2n}}^{\widehat{\theta}} & . & . & . & {Z_{pn}}^{\widehat{\theta}}
\end{bmatrix} = {Z_{iu}}^{\widehat{\theta}}$$

Where $\widehat{Z}'$ is

$$Z'_{\widehat{\theta}} = \begin{bmatrix} {Z_{11}}^{\widehat{\theta}} & {Z_{12}}^{\widehat{\theta}} & . & . & . & {Z_{1n}}^{\widehat{\theta}} \\ {Z_{21}}^{\widehat{\theta}} & {Z_{22}}^{\widehat{\theta}} & . & . & . & {Z_{2n}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{u1}}^{\widehat{\theta}} & {Z_{u2}}^{\widehat{\theta}} & . & . & . & {Z_{up}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{p1}}^{\widehat{\theta}} & {Z_{p2}}^{\widehat{\theta}} & . & . & . & {Z_{pn}}^{\widehat{\theta}} \end{bmatrix} = ({Z_{iu}}^{\widehat{\theta}})'$$

So that $\widehat{Z}'\widehat{Z}$ is

$$Z'_{\widehat{\theta}} Z_{\widehat{\theta}} = \begin{bmatrix} {Z_{11}}^{\widehat{\theta}} & {Z_{12}}^{\widehat{\theta}} & . & . & . & {Z_{1n}}^{\widehat{\theta}} \\ {Z_{21}}^{\widehat{\theta}} & {Z_{22}}^{\widehat{\theta}} & . & . & . & {Z_{2n}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{u1}}^{\widehat{\theta}} & {Z_{u2}}^{\widehat{\theta}} & . & . & . & {Z_{up}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{p1}}^{\widehat{\theta}} & {Z_{p2}}^{\widehat{\theta}} & . & . & . & {Z_{pn}}^{\widehat{\theta}} \end{bmatrix} \begin{bmatrix} {Z_{11}}^{\widehat{\theta}} & {Z_{21}}^{\widehat{\theta}} & . & . & . & {Z_{p1}}^{\widehat{\theta}} \\ {Z_{12}}^{\widehat{\theta}} & {Z_{22}}^{\widehat{\theta}} & . & . & . & {Z_{p2}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{1u}}^{\widehat{\theta}} & {Z_{2u}}^{\widehat{\theta}} & . & . & . & {Z_{pu}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{1n}}^{\widehat{\theta}} & {Z_{2n}}^{\widehat{\theta}} & . & . & . & {Z_{pn}}^{\widehat{\theta}} \end{bmatrix} = ({Z_{iu}}^{\widehat{\theta}})' {Z_{iu}}^{\widehat{\theta}}$$

SE_{yx} is Standard error of the estimate as below

$$SE_{yx} = \sqrt{\frac{SSE}{n-d-1}} = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n-d-1}}$$

Where

$$SSE = \sum_{i=1}^{n}(\widehat{y_i} - y_i)^2$$

So that $\mathrm{se}(\widehat{\theta})$ is as below

$$se(\widehat{\theta}) = \left[({Z_{iu}}^{\widehat{\theta}})' {Z_{iu}}^{\widehat{\theta}}\right]^{-1/2} \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n-d-1}$$

For the coefficient θ include s_∞, σ, w, k, z. In this case, for the coefficient s_∞, σ, w, k, z we got the Standard Error (SE) of an estimate of each coefficient as below.

$$SE_{\widehat{s_\infty}} = \left[({Z_{iu}}^{\widehat{s_\infty}})' {Z_{iu}}^{\widehat{s_\infty}}\right]^{-1/2} \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n-d-1}$$

$$SE_{\widehat{\sigma}} = \left[(Z_{iu}{}^{\widehat{\sigma}})' Z_{iu}{}^{\widehat{\sigma}}\right]^{-1/2} \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n-d-1}$$

$$SE_{\widehat{w}} = \left[(Z_{iu}{}^{\widehat{w}})' Z_{iu}{}^{\widehat{w}}\right]^{-1/2} \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n-d-1}$$

$$SE_{\widehat{k}} = \left[(Z_{iu}{}^{\widehat{k}})' Z_{iu}{}^{\widehat{k}}\right]^{-1/2} \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n-d-1}$$

$$SE_{\widehat{z}} = \left[(Z_{iu}{}^{\widehat{z}})' Z_{iu}{}^{\widehat{z}}\right]^{-1/2} \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n-d-1}$$

Coefficient of Determination (R^2) of the Island Biogeographic Model base on Nonlinear Normal Regression

The coefficient of determination (R^2) is an important statistic reflecting how to estimate the model correctly, and it is the ratio of the regression sum of squares to the total sum of squares.

The value of R^2 takes a value between 0 and 1, and has no unit. Its value reflects the relative degree of regression contribution, that is the percentage that the regression relationship can explain in the total variation of the dependent variable Y. R^2 is the most commonly used index to evaluate the pros and cons of regression models. The larger the R^2 (closer to 1) is, it fit the better the regression equation.

The value of R^2 used in Nonlinear Normal Regression. The way we used to measure is the pseudo R^2, defined as.

$$R^2 = \frac{\ell(\theta \mid \widehat{X_0}, Y) - \ell(\theta \mid X, Y)}{\ell(\theta \mid \widehat{X_0}, Y)} = 1 - \frac{\ell(\theta \mid X, Y)}{\ell(\theta \mid \widehat{X_0}, Y)}$$

Where ℓ is given by

$$\ell(\theta \mid X, Y) = \log L(\theta \mid X, Y) = -\frac{nlog(2\pi)}{2} - \frac{nlog(\sigma^2)}{2} - \frac{1}{(2\sigma^2)} \sum_{i=0}^{n} (y_i - s_\infty(1 + (we^{zIna} - 1)e^{-kt})$$

$\ell(\theta \mid \widehat{X_0}, Y)$ is the log likelihood of the model from Nonlinear Normal Regression. The pseudo R^2 goes from 0 to 1, with 1 being the perfect fit.

Errors and Residual of the Island Biogeographic Model base on the Nonlinear Normal Regression

In statistics and optimization, errors and residuals are too closely related, and easily confusing measures of the deviation of an observed value of an item, in a statistical sample from its theoretical value. The error of an observed value is the deviation of the observed value from the true value of a quantity of interest, and the residual of an observed value is the difference between the observed value and the estimated value of the quantity of interest.

From the Island Biogeographic Model date, we can define Y as $S(t)$, and the arrangement is

$$S(t)_i \sim S(t)_1, \ldots, S(t)_n$$

Where residual r_i is

$$r_i = S(t)_i - \hat{S}(t)_i = s_\infty(1 + (we^{zIna} - 1)e^{-kt}) - \hat{S}(t)_i$$

The residual standard error calculated as

$$RME = \sqrt{\sum_{i=1}^{n} \frac{r^2}{df}} = \sqrt{\frac{1}{n-d-1} \sum_{i=1}^{n} (S(t)_i - \hat{S}(t)_i)^2}$$

Where r is the residual and df the residual degrees of freedom(n minus d), where the sample size is n and the estimated parameters are d.

Akaike Information Criterion (AIC) of the Island Biogeographic Model based on the Nonlinear Normal Regression

Akaike information criterion (AIC) is a standard for measuring the suitability of statistical model fitting. The Akaike information criterion is based on the concept of entropy, which can weight the complexity of the estimated model and the suitability of the model to fit the data. The lower the AIC is, the more concise and accurate the fitting model will be.
From the Island Biogeographic Model with Nonlinear Normal Regression, it can express AIC as it follows:

$$AIC = 2d - 2log\mathcal{L}$$

Among them, according to the above which d is the number of parameters, $\mathcal{L}$ is the likelihood function according to the Nonlinear Normal Regression as below:

$$log\mathcal{L}(x_1, \ldots, x_n \mid \theta, \sigma) = \frac{-nlog(2\pi)}{2} - \frac{nlog(\sigma^2)}{2} - \frac{1}{(2\sigma^2)} \sum_{i=0}^{n} (y_i - s_\infty(1 + (we^{zIna} - 1)e^{-kt})$$

Above that show that the model errors follow a Nonlinear Normal Regression. n is the number of observations, SS_{res} is the residual sum of squares, then AIC becomes as you can see below:

$$AIC = 2d + nlog(\frac{SS_{\text{res}}}{n})$$

Where the residual sum of squares is

$$SS_{\text{res}} = \sum_{i=0}^{n} (y_i - s_\infty(1 + (we^{zIna} - 1)e^{-kt})^2 = \sum_{i=0}^{n} e_i^2$$

2.4. Concept based on Nonlinear Poisson Regression

For Nonlinear Poisson Regression Models, we can get the following relation:

$$Y_i \sim P(\lambda(x, \theta), \sigma^2)$$

Y_i is the vector of response variable, x_i is the vector of explanatory variable, θ are the parameters vectors of s_∞, σ, w, k and z.
We insert the formula into Nonlinear Poisson Regression, for the Nonlinear Poisson distribution $\mathcal{P}(\lambda(x, \theta), \sigma^2)$ having density function.
Given the θ and explanatory variables $\mathbf{a}, \mathbf{t}$ in the Nonlinear Poisson Regression model, the regression expression is

$$\begin{aligned} \lambda = E(Y \mid a, t) &= s_\infty - s_\infty e^{-kt} + ca^z e^{-kt} \\ &= s_\infty(1 + (we^{zIna} - 1)e^{-kt}) \end{aligned}$$

And thus, the Poisson Distribution's Function is given by

$$f(y \mid a, t; \theta) = \frac{\lambda^y}{y!} e^{-\lambda} = \frac{(s_\infty(1 + (we^{zIna} - 1)e^{-kt}))^y}{y!} e^{-(s_\infty(1+(we^{zIna}-1)e^{-kt}))}$$

It is now known that the observed value of the explanatory variable is composed of n vectors $a_i, t_i \in \mathbb{R}^{n+1}, i = 1, \ldots, n$ corresponding to the observations of n explained variables, $y_1, \ldots, y_n \in \mathbb{R}$ where $\theta = (s_\infty, \sigma, w, k, z)$ because λ is longer and hardly to display, so we put λ as

$$\lambda = E(Y \mid a, t) = s_\infty - s_\infty e^{-kt} + ca^z e^{-kt}$$

Then the set of observations corresponding to the joint function can be expressed by the following formula

$$\begin{aligned} f(y_1, \ldots, y_m \mid a_1, \ldots, a_n, t_1, \ldots, t_n; \theta) &= \prod_{i=1}^{n} \frac{(\lambda)^y e^{-(\lambda)}}{y!} \\ &= \prod_{i=1}^{n} \frac{(s_\infty(1 + (we^{zIna} - 1)e^{-kt}))^y}{y!} e^{-(s_\infty(1+(we^{zIna}-1)e^{-kt}))} \end{aligned}$$

The core idea of the maximum likelihood method for estimating θ is to find the θ maximizing the joint function, based on the current observations. Since the goal is to find the optimal θ, we can get X_i and include a_i, t_i. It can simply express the left side of the equal sign as an expression about θ

$$L(\theta \mid X, Y) = \prod_{i=1}^{n} \frac{(\lambda)^y e^{-(\lambda)}}{y!} = \prod_{i=1}^{n} \frac{(s_\infty(1 + (we^{zIna} - 1)e^{-kt}))^y}{y!} e^{-(s_\infty(1+(we^{zIna}-1)e^{-kt}))}$$

The expression on the right side of the equal sign is not rewritten, but it is difficult to calculate, so the log-likelihood expression is used as it follows:

$$\begin{aligned}\ell(\theta \mid X,Y) &= \log L(\theta \mid X,Y)\\ &= \sum_{i=1}^{n}(y_i log\lambda - \lambda - \log(y_i!))\\ &= \sum_{i=1}^{n}\big(y_i log(s_\infty(1+(we^{zIna}-1)e^{-kt})) - (s_\infty(1+(we^{zIna}-1)e^{-kt})) - \log(y_i!)\big)\end{aligned}$$

Since θ only appears in the first two terms of the likelihood function, in the process of maximizing the likelihood function, only the first two terms can be considered. The third term $y_i!$ can be deleted, and the likelihood function to be optimized can be simply expressed as it follows:

$$\begin{aligned}\ell(\theta \mid X,Y) = \log L(\theta \mid X,Y) &= \sum_{i=1}^{n}(y_i log\lambda - \lambda)\\ &= \sum_{i=1}^{n}\big(y_i log(s_\infty(1+(we^{zIna}-1)e^{-kt})) - (s_\infty(1+(we^{zIna}-1)e^{-kt}))\big)\end{aligned}$$

Maximum Likelihood Estimation of the Island Biogeographic Model based on Nonlinear Poisson Regression

Maximum likelihood estimation (MLE) is an important and common method for obtaining estimators. The maximum likelihood method explicitly uses a model, and its goal is to find a phylogenetic tree being able to produce observation data with a higher probability.
In order to find the maximum value, the following equation needs to be solved:

$$\frac{\partial \ell(\theta \mid X,Y)}{\partial \theta} = 0$$

To maximize its logarithms need to maximize for relation of θ, because θ is the vector parameters included s_∞, σ, w, k, z, so that we need to separate to each parameter to find the relationship for each parameter.
To Maximize for s_∞, the partial derivative of the logarithms likelihood must be equal to zero, regarding s_∞, we got the formula as below:

$$\frac{\partial \ell(\theta \mid X,Y)}{\partial s_\infty} = 0$$

$$\sum_{i=1}^{n}(y_i log'(s_\infty(1+(we^{zIna}-1)e^{-kt}))(1+(we^{zIna}-1)e^{-kt}) - \sum_{i=1}^{n}(1+(we^{zIna}-1)e^{-kt}) = 0$$

$$\sum_{i=1}^{n}(y_i log'(s_\infty(1+(we^{zIna}-1)e^{-kt}) - 1)(1+(we^{zIna}-1)e^{-kt}) = 0$$

For this part we could find out that for $(y_i log'(s_\infty(1+(we^{zIna}-1)e^{-kt})-1)(1+(we^{zIna}-1)e^{-kt})$ is zero, here we could not find out the relationship only with formula, so for this part we need to use R language program to find out the value of each parameter.
To Maximize for w, the partial derivative of the logarithms likelihood must be equal to zero, regarding w we got the formula as below:

$$\frac{\partial \ell(\theta \mid X,Y)}{\partial w} = 0$$

$$\sum_{i=1}^{n}(y_i log'(s_\infty(1+(we^{zIna}-1)e^{-kt}))((s_\infty e^{zIna})e^{-kt}) - \sum_{i=1}^{n}(s_\infty e^{zIna})e^{-kt})) = 0$$

$$\sum_{i=1}^{n}(y_i log'(s_\infty(1+(we^{zIna}-1)e^{-kt}))-1))(s_\infty e^{zIna}e^{-kt}) = 0$$

For this part we could find out that for $(y_i log'(s_\infty(1+(we^{zIna}-1)e^{-kt}))-1))(s_\infty e^{zIna}e^{-kt})$ is zero, here we could not find out the relationship only with formula, so for this part we need to use R language program to find out the value of each parameter.
To Maximize for z, the partial derivative of the logarithms likelihood must be equal to zero, regarding z we got the formula as below:

$$\frac{\partial \ell(\theta \mid X,Y)}{\partial z} = 0$$

$$\sum_{i=1}^{n}(y_i log'(s_\infty(1+(we^{zIna}-1)e^{-kt}))((s_\infty e^{zIna})e^{-kt}) - \sum_{i=1}^{n}(s_\infty e^{zIna})e^{-kt})) = 0$$

$$\sum_{i=1}^{n}(y_i log'(s_\infty(1+(we^{zIna}-1)e^{-kt}))-1))(s_\infty e^{zIna}e^{-kt}) = 0$$

For this part we could find out that for $(y_i log'(s_\infty(1+(we^{zIna}-1)e^{-kt}))-1))(s_\infty e^{zIna}e^{-kt})$ is zero, here we could not find out the relationship only with formula, so for this part we need to use R language program to find out the value of each parameter.
To Maximize for k, the partial derivative of the logarithms likelihood must be equal to zero, regarding k we got the formula as below:

$$\frac{\partial \ell(\theta \mid X,Y)}{\partial k} = 0$$

$$\sum_{i=1}^{n}(y_i log'(s_\infty(1+(we^{zIna}-1)e^{-kt}))(-s_\infty(we^{zIna}-1)e^{-kt}) - \sum_{i=1}^{n}(-s_\infty(we^{zIna}-1)e^{-kt}) = 0$$

$$\sum_{i=1}^{n}(1-y_i log'(s_\infty(1+(we^{zIna}-1)e^{-kt})))(s_\infty(we^{zIna}-1)e^{-kt}) = 0$$

For this part we could find out that for $(1 - y_i log'(s_\infty(1 + (we^{zIna} - 1)e^{-kt})))(s_\infty(we^{zIna} - 1)e^{-kt})$ is zero, here we could not find out the relationship only with formula, so for this part we need to use R language program to find out the value of each parameter.

Where $\theta = (s_\infty, w, k, z)$ has no closed-form solution. however, the negative log-likelihood, $-\ell(\theta \mid X, Y)$ is a convex function, and includes convex optimization techniques such as gradient descent, applied to find the optimal value of θ. from the above formula, the corresponding relationship of each parameter can be got. Here, we need to use the R language program to estimate each parameter.When using the R language program, we need to use the nls and gnm packages, take the formula from section 2.2 and choose the Nonlinear Poisson Regression format, import the data, and draw conclusions based on the analysis of the data to estimate the value of each parameter.

Confidence Intervals for Regression Coefficients of the Island Biogeographic Model based on the Nonlinear Poisson Regression

Allaire et al. (2021) showed that the problem to be solved by the coefficient confidence interval is to examine how close the estimated coefficient value is, to the true value of the coefficient. The smaller the coefficient confidence interval is, the more effective expression of the coefficient will be. For nonlinear Poisson regression fitting, the coefficient estimation value has been got from section 2.4, which is the value of parameters vectors $\widehat{\theta}$ included $\widehat{s}_\infty$, $\widehat{\sigma}$, $\widehat{c}$, $\widehat{k}$ and $\widehat{z}$. for the coefficient confidence interval of nonlinear Poisson regression, we get the formula as it follows

$$\mathrm{CI}^{\hat{\theta}}_{0,95} = \left[\hat{\theta} - t_{\alpha/2}(n - d - 1) \times SE_{\widehat{\theta}}\,,\, \hat{\theta} + t_{\alpha/2}(n - d - 1) \times SE_{\widehat{\theta}}\right].$$

In generally the number of parameters we defined as k, but in the regression there have the coefficient k, to avoid the difference between number of parameters k and parameter of $\widehat{k}$, we defined the number of parameters as d. α is 5 %, from t table we got the value of t base on $\alpha/2$ is 0.025, from the date we got that n is 64, d is 4, $n-d-1$ is 59, so that $t_{\alpha/2}(n - d - 1)$ is 2.00. The Standard Error (SE) of an estimate of a coefficient is the Standard Deviation of the estimate.

What we can get λ as

$$\begin{aligned} \lambda = \mathrm{E}(Y \mid a, t) &= s_\infty - s_\infty e^{-kt} + ca^z e^{-kt} \\ &= s_\infty(1 + (we^{zIna} - 1)e^{-kt}) \end{aligned}$$

So that we can get the coefficient confidence interval of the vectors $\widehat{\theta}$ included $\widehat{s}_\infty$, $\widehat{\sigma}$, $\widehat{c}$, $\widehat{k}$

and $\widehat{z}$, it is showed as below

$$\begin{aligned}
\mathrm{CI}_{0,95}^{\widehat{s_\infty}} &= \left[\hat{s}_\infty - t_{\alpha/2}(n-d-1)\times SE_{\widehat{s_\infty}}\,,\ \hat{s}_\infty + t_{\alpha/2}(n-d-1)\times SE_{\widehat{s_\infty}}\right].\\
\mathrm{CI}_{0,95}^{\widehat{\sigma}} &= \left[\hat{\sigma} - t_{\alpha/2}(n-d-1)\times SE_{\widehat{\sigma}}\,,\ \hat{\sigma} + t_{\alpha/2}(n-d-1)\times SE_{\widehat{\sigma}}\right].\\
\mathrm{CI}_{0,95}^{\widehat{c}} &= \left[\hat{c} - t_{\alpha/2}(n-d-1)\times SE_{\widehat{c}}\,,\ \hat{c} + t_{\alpha/2}(n-d-1)\times SE_{\widehat{c}}\right].\\
\mathrm{CI}_{0,95}^{\widehat{k}} &= \left[\hat{k} - t_{\alpha/2}(n-d-1)\times SE_{\widehat{k}}\,,\ \hat{k} + t_{\alpha/2}(n-d-1)\times SE_{\widehat{k}}\right].\\
\mathrm{CI}_{0,95}^{\widehat{z}} &= \left[\hat{z} - t_{\alpha/2}(n-d-1)\times SE_{\widehat{z}}\,,\ \hat{z} + t_{\alpha/2}(n-d-1)\times SE_{\widehat{z}}\right].
\end{aligned}$$

When use Nonlinear Poisson Regression to estimate regression analysis on each column from the data, for each coefficient aspect to each variable, so here it can concluded that the defined the Standard Error (SE) of an estimate of each coefficient, according to Draper Norman et al. (1998)(chapter 24) we got the definition for Nonlinear Poisson Regression the standard error of the regression slope coefficient θ_i is estimated as se(θ) = {appropriate diagonal element of $((\widehat{Z}'\widehat{Z})^{-1}SE_{yx}^2)^{1/2}$ }.
Where ${Z_{iu}}^{\widehat{\theta}}$ is to maximize its logarithms need to maximize for relation of θ as below

$${Z_{iu}}^{\widehat{\theta}} = \left[\frac{\partial \ell(\theta \mid X, Y)}{\partial \theta}\right]_\theta$$

Where $\widehat{Z}$ is the Z of estimated value

$$Z_{\widehat{\theta}} = \begin{bmatrix}
{Z_{11}}^{\widehat{\theta}} & {Z_{21}}^{\widehat{\theta}} & . & . & . & {Z_{p1}}^{\widehat{\theta}}\\
{Z_{12}}^{\widehat{\theta}} & {Z_{22}}^{\widehat{\theta}} & . & . & . & {Z_{p2}}^{\widehat{\theta}}\\
. & . & & & & .\\
. & . & & & & .\\
. & . & & & & .\\
{Z_{1u}}^{\widehat{\theta}} & {Z_{2u}}^{\widehat{\theta}} & . & . & . & {Z_{pu}}^{\widehat{\theta}}\\
. & . & & & & .\\
. & . & & & & .\\
. & . & & & & .\\
{Z_{1n}}^{\widehat{\theta}} & {Z_{2n}}^{\widehat{\theta}} & . & . & . & {Z_{pn}}^{\widehat{\theta}}
\end{bmatrix} = {Z_{iu}}^{\widehat{\theta}}$$

Where $\widehat{Z}'$ is

$$Z'_{\widehat{\theta}} = \begin{bmatrix} {Z_{11}}^{\widehat{\theta}} & {Z_{12}}^{\widehat{\theta}} & . & . & . & {Z_{1n}}^{\widehat{\theta}} \\ {Z_{21}}^{\widehat{\theta}} & {Z_{22}}^{\widehat{\theta}} & . & . & . & {Z_{2n}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{u1}}^{\widehat{\theta}} & {Z_{u2}}^{\widehat{\theta}} & . & . & . & {Z_{up}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{p1}}^{\widehat{\theta}} & {Z_{p2}}^{\widehat{\theta}} & . & . & . & {Z_{pn}}^{\widehat{\theta}} \end{bmatrix} = ({Z_{iu}}^{\widehat{\theta}})'$$

So that $\widehat{Z}'\widehat{Z}$ is

$$Z'_{\widehat{\theta}} Z_{\widehat{\theta}} = \begin{bmatrix} {Z_{11}}^{\widehat{\theta}} & {Z_{12}}^{\widehat{\theta}} & . & . & . & {Z_{1n}}^{\widehat{\theta}} \\ {Z_{21}}^{\widehat{\theta}} & {Z_{22}}^{\widehat{\theta}} & . & . & . & {Z_{2n}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{u1}}^{\widehat{\theta}} & {Z_{u2}}^{\widehat{\theta}} & . & . & . & {Z_{up}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{p1}}^{\widehat{\theta}} & {Z_{p2}}^{\widehat{\theta}} & . & . & . & {Z_{pn}}^{\widehat{\theta}} \end{bmatrix} \begin{bmatrix} {Z_{11}}^{\widehat{\theta}} & {Z_{21}}^{\widehat{\theta}} & . & . & . & {Z_{p1}}^{\widehat{\theta}} \\ {Z_{12}}^{\widehat{\theta}} & {Z_{22}}^{\widehat{\theta}} & . & . & . & {Z_{p2}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{1u}}^{\widehat{\theta}} & {Z_{2u}}^{\widehat{\theta}} & . & . & . & {Z_{pu}}^{\widehat{\theta}} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ {Z_{1n}}^{\widehat{\theta}} & {Z_{2n}}^{\widehat{\theta}} & . & . & . & {Z_{pn}}^{\widehat{\theta}} \end{bmatrix} = ({Z_{iu}}^{\widehat{\theta}})' {Z_{iu}}^{\widehat{\theta}}$$

SE_{yx} is Standard error of the estimate as below

$$SE_{yx} = \sqrt{\frac{SSE}{n-d-1}} = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n-d-1}}$$

Where

$$SSE = \sum_{i=1}^{n}(\widehat{y_i} - y_i)^2$$

So that $se(\widehat{\theta})$ is as below

$$se(\widehat{\theta}) = [({Z_{iu}}^{\widehat{\theta}})' {Z_{iu}}^{\widehat{\theta}}]^{-\frac{1}{2}} \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n-d-1}$$

Here coefficient θ include s_∞, σ, w, k, z, in this case for the coefficient s_∞, σ, w, k, z we got the Standard Error (SE) of an estimate of each coefficient as below

$$SE_{\widehat{s_\infty}} = \left[({Z_{iu}}^{\widehat{s_\infty}})' {Z_{iu}}^{\widehat{s_\infty}}\right]^{-\frac{1}{2}} \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n-d-1}$$

$$SE_{\widehat{\sigma}} = \left[(Z_{iu}{}^{\widehat{\sigma}})' Z_{iu}{}^{\widehat{\sigma}} \right]^{-\frac{1}{2}} \frac{\sum_{i=1}^{n} (\widehat{y}_i - y_i)^2}{n-d-1}$$

$$SE_{\widehat{w}} = \left[(Z_{iu}{}^{\widehat{w}})' Z_{iu}{}^{\widehat{w}} \right]^{-\frac{1}{2}} \frac{\sum_{i=1}^{n} (\widehat{y}_i - y_i)^2}{n-d-1}$$

$$SE_{\widehat{k}} = \left[(Z_{iu}{}^{\widehat{k}})' Z_{iu}{}^{\widehat{k}} \right]^{-\frac{1}{2}} \frac{\sum_{i=1}^{n} (\widehat{y}_i - y_i)^2}{n-d-1}$$

$$SE_{\widehat{z}} = \left[(Z_{iu}{}^{\widehat{z}})' Z_{iu}{}^{\widehat{z}} \right]^{-\frac{1}{2}} \frac{\sum_{i=1}^{n} (\widehat{y}_i - y_i)^2}{n-d-1}$$

Coefficient of Determination (R^2) of the Island Biogeographic Model based on Nonlinear Poisson Regression

The value of R^2 is used in Nonlinear Poisson Regression. The way we used to measure is the pseudo R^2 defined as

$$R^2 = \frac{\ell(\theta \mid \widehat{X_0}, Y) - \ell(\theta \mid X, Y)}{\ell(\theta \mid \widehat{X_0}, Y)} = 1 - \frac{\ell(\theta \mid X, Y)}{\ell(\theta \mid \widehat{X_0}, Y)}$$

Where ℓ is given by

$$\ell(\theta \mid X, Y) = \log L(\theta \mid X, Y) = \sum_{i=1}^{n} (y_i log(s_\infty(1 + (we^{zIna} - 1)e^{-kt})) - (s_\infty(1 + (we^{zIna} - 1)e^{-kt})))$$

$\ell(\theta \mid \widehat{X_0}, Y)$ is the log likelihood of the model from Nonlinear Poisson Regression. The pseudo R^2 goes from 0 to 1, where 1 is the perfect fit.

Errors and Residuals of the Island Biogeographic Model based on Nonlinear Poisson Regression

From the Island Biogeographic Model date, we can define Y as $S(t)$, and the arrangement is

$$S(t)_i \sim S(t)_1, \ldots, S(t)_n$$

where residual r_i is

$$r_i = S(t)_i - \hat{S}(t)_i = s_\infty(1 + (we^{zIna} - 1)e^{-kt}) - \hat{S}(t)_i$$

We calculated the residual standard error as it follows:

$$RME = \sqrt{\sum_{i=1}^{n} \frac{r^2}{df}} = \sqrt{\frac{1}{n-d-1} \sum_{i=1}^{n} (S(t)_i - \hat{S}(t)_i)^2}$$

Where r is the residual and df the residual degrees of freedom(n minus d), where the sample size is n and the number of estimated parameter is d.

Akaike Information Criterion (AIC) of the Island Biogeographic Model base on Nonlinear Poisson Regression

From the Nonlinear Poisson Regression, we can get the following results $Y_i \sim P(\lambda(x,\theta),\sigma^2,\ell)$, where $\ell = (\ell_1,...,\ell_N)^T$ is the mean of the Poisson distribution and depends on some covariance and the random effects ℓ. The Island Biogeographic Model modeled the data with conditional likelihood denoted by $L(Y \mid X,\theta)$. For any estimator $\widehat{X}_{(Y)}$ and $\widehat{\theta}_{(Y)}$, an unbiased estimator of the AIC is:

$$AIC = -2logL(Y \mid X,\theta) + 2K$$

Where K is given by:

$$\sum_{i=1}^{N}\{y_i log\{\widehat{\ell}_i(y)\} - y_i log\{\widehat{\ell}_i(y^{y_i-1})\}\}$$

Where ℓ is given by:

$$\ell_i(y) = \log L(\theta \mid X,Y) = \sum_{i=1}^{n}\left(y_i log(s_\infty(1+(we^{zIna}-1)e^{-kt})) - (s_\infty(1+(we^{zIna}-1)e^{-kt}))\right)$$

$\widehat{\ell}_i(y)$ is the fitted value of ℓ_i based on data y, y^{y_i-1} is the same as y except its i-th component is replaced by $y_i - 1$, and in another hand, when $y_i = 0$, we can got $y_i log\{\widehat{\ell}_i(y^{y_i-1})\} = 0$.

3 Objectives

3.1. General Objectives

Develop the estimation theory of The Nonlinear Poisson Regression model for species extinction in fragmented habitats.

3.2. Specific Objectives

1. Analyze and comment on the results between the Nonlinear Poisson Regression Model and the Nonlinear Normal Regression Model.
2. Perform Maximum Likelihood Estimation (MLE) with the obtained data, and analyze the different conclusions.
3. Calculate the average time when the species extinction approaches to the half.

4 Methodology

We will analyze the all information through the R Program Language version 4.0.2.

4.1. The Manner of R Program Language

It is necessary to use Nonlinear Normal Regression and Nonlinear Poisson Regression to realize the regression analysis of the data. When using the R language programming, First is to determine the data that needs to import, and then confirm the species formula. in this case, we need to use the formula from section 2.2, and then bring in the required program package. In this case, we use *nls* and *gnm* packages. Write and organize the code, and then running the code to get the results of Nonlinear Normal Regression and Nonlinear Poisson Regression respectively, and measure the value of R^2 and Aic and $residual$.

4.2. MLE Method

In order to get the results, we need to use the R program language to fit the data, and the method used for the fitting is MLE.

4.3. Graphical Method

In order to plot the changes in the number of remaining species in each island with time passing, and the remaining half of the species extinct, it is also necessary to use the R program language. Based on the date estimate through R program language to The Nonlinear Normal Regression and The Nonlinear Poisson Regression to find the value of each parameter, and then it is estimated that in the time period of 7 - 80 years. Here the island areas are 1 ha, 5 ha, 10 ha, 25 ha, 50 ha, the number of remaining species on the respective islands with time passing, with Nonlinear Normal Regression and Nonlinear Poisson Regression, and the Regression from Luke et al. (2013), then calculate the extinction rate corresponding to these three regressions, and then find out the average value of the islands in the data, and the Nonlinear Normal Regression and Nonlinear Poisson Regression and the Regression from Luke et al. (2013) were used respectively to calculate the number of remaining species in this time period from 7 to 80 years, and then find out the time when the remaining species in different island areas become extinct to half of the three regression.

4.4. Analysis of the Information

All the above articles have shown that human activities have had a serious impact on the destruction of nature, and the statistical methods used in most of the articles are Multivariate Nonlinear Normal Regression Models, and papers by Mendes et al. (2016), Megan et al. (2011), Martensen et al. (2012), Chisholm et al. (2018) use the Nonlinear Poisson Regression model to estimate the data, so it can show more effective conclusions.

The Nonlinear Poisson Regression model and Multivariate Nonlinear Normal Regression Model are suitable for the calculation of data related to species extinction, and the Nonlinear Poisson Regression Model is more suitable for calculations, including time variables.

The same method applies to calculate the following data from southern Thailand. We applied this method as it is described as below:

1. Use the Formulas and Nonlinear Normal Regression Models to estimate the data.
2. Use the Formulas and Nonlinear Poisson Regression Model to estimate the data.
3. Make comparisons between conclusions and finally show which model is more effective.

5 Result

No.	Original	Nonlinear Normal	Nonlinear Poisson
s_∞	0.7510	0.9640	1.2192
$w(c/s_\infty)$	2.9600	2.3081	1.8288
z	0.4820	0.8719	0.8803
k	0.0732	0.0732	0.0780
c	2.2229	2.2251	2.2297
R^2	0.7830	0.8660	0.8660
AIC		369.4768	259.4849
Residual deviance		1014/60 degrees	60.6040/60 degrees

Table 5-1: Results of Nonlinear Normal Regression and Nonlinear Poisson Regression and Original Data Fitting

According to Luke et al. (2013), the corresponding relational formula is deduced as it follows:

$$S(t) = s_\infty - (s_\infty - ca^z)e^{-kt}$$

Then we get the result from the original article,the formula can be showed as below:

$$S(t) = 0{,}7510 - (0{,}7510 - 2{,}2230a^{0{,}4820})e^{-0{,}0732t}$$

Our purpose is recalculating a regression model that is more effective with the mathematical standards and thinking about how to use this data to further bring out the processing methods for this type of data. In Luke et al. (2013), they used a Nonlinear Normal Regression Model,and the formula used was.

$$S(t) = s_\infty - (s_\infty - ca^z)e^{-kt}$$

In order to make the calculation more convenient, the formula needs to be redefined.
From the original formula as below

$$S(t) = s_\infty - (s_\infty - ca^z)e^{-kt}$$

Then we need to redefine the formula as we show it as below (Section 2.2)

$$S(t) = s_\infty(1 + (we^{zIna} - 1)e^{-kt})$$

Where $w = c/s_\infty$.

Result from Original Luke et al. (2013)

The data got in Luke et al. (2013) is displayed in the new formula as it follows

$$S(t) = s_\infty(1 + (we^{zIna} - 1)e^{-kt})$$

Where $w = c/s_\infty$

$$S(t) = 0{,}7510(1 + (2{,}9600e^{0{,}4820Ina} - 1)e^{-0{,}0732t})$$

Where the parameter s_∞ is 0.7510, $w(c/s_\infty)$ is 2.9600, z is 0.4820, k is 0.0732, c is 2.2229. The coefficient of determination is 0.7830. Because it is the original conclusion, there is no calculation of AIC and Residual Deviance. The value of coefficient of determination is 0.7830 which means that 78.3 % of the data can be effectively shown in this Nonlinear Normal Regression.

According to the values of the above parameters, we found that the number of remaining species has a positive correlation with the area of the island, the larger the area of the island is, the higher the number of residues of species on the island is. However, there is a negative correlation between time and the number of residues of species. The passing of the years makes the number of remaining species gradually decrease, and little by little species disappear.

Result from Nonlinear Normal Regression

In comparison, then we estimate Nonlinear Normal Regression on the data, and the conclusion is as below:

$$S(t) = s_\infty(1 + (we^{zIna} - 1)e^{-kt})$$

Where $w = c/s_\infty$.

$$S(t) = 0{,}9640(1 + (2{,}3081e^{0{,}8719Ina} - 1)e^{-0{,}0732t})$$

Where parameter s_∞ is 0.9640, $w(c/s_\infty)$ is 2.3081, z is 0.8719, k is 0.0732, c is 2.2251. The Coefficient of Determination is 0.8660 and AIC is 369.4768. The value of Coefficient of Determination is 0.8660, it means that 86,60 % of the data can be effectively displayed in this Nonlinear Normal Regression. The value of Akaike Information Criterion (AIC) is 369.4768. The Residual Deviance is 1014 on 60 degrees of freedom.

According to the values of the above parameters, we found that the number of remaining species has a positive correlation with the area of the island. The larger the size of the island is, the larger is the number of residues on the island. However, there is a negative correlation between time and the quantity remaining. The passing of the years makes the number of remaining species gradually decrease, and the species gradually become extinct.

Compared with the previous parameter, in the data of Luke et al. (2013), the value of c is 2.2229. In the new Nonlinear Normal Regression, the value of c is 2.2251, and the value of z

from Luke et al. (2013) is smaller than the value in the new Nonlinear Normal Regression, this can be explained that when the area of the island increases, the increase in the number of remaining species will be higher than the model in Luke et al. (2013) using the Nonlinear Normal Regression.
The value of k is similar on both sides, but because of the difference in other parameters, it can also be concluded that with the passing of time, the rate of species extinction of the Nonlinear Normal Regression Model is faster than the model in Luke et al. (2013).
From Table 2 the results of the parameters Confidence Interval of regression of Nonlinear Normal parameters are 2.5 % - 97.5 % respectively s_∞ is from -1.4691 to 3.0141, w is from 1.0608 to 3.8145, z is from 0.7296 to 1.0620, k is from 0.0545 to 0.0998.
For the Parameter Confidence Interval, the problem to be solved is to examine how close the estimated parameter value is to the true value of the parameter. For the parameter value of the Nonlinear Normal Regression obtained by the fitting , where the parameter s_∞ is 0.964, $w(c/s_\infty)$ is 2.3081, z is 0.8719, k is 0.0732, c is 2.2251. Comparing the corresponding parameter confidence interval, we can establish that all the fitted values are in the parameter confidence interval.
The data shows that the Nonlinear Normal Regression for the data obtained by using *nls* packages and *gnm* packages is the same, showing that if the Nonlinear Normal Regression is used, because the coefficient of determination is higher than the original, and the AIC is more or less as the original, the conclusion is more effective than the conclusion drawn in Luke et al. (2013).

Result from Nonlinear Poisson Regression

For the data, all the parameters which include time and area, these parameters are effective for Poisson Regression, so Nonlinear Poisson Regression is used for analysis. We estimate to perform Nonlinear Poisson Regression on the data, and the conclusion is as it follows

$$S(t) = s_\infty(1 + (we^{zIna} - 1)e^{-kt}))$$

Where $w = c/s_\infty$

$$S(t) = 1{,}2192(1 + (1{,}8288e^{0{,}8803Ina} - 1)e^{-0{,}0780t}))$$

Where s_∞ is 1.2192, $w(c/s_\infty)$ is 1.8288, z is 0.8803, k is 0.0780, c is 2.22972, the coefficient of determination is 0.8660 and AIC is 259.4849. The value of coefficient of determination is 0.8660. It means that 86,6 % of the data can be effectively displayed in this Nonlinear Poisson Regression. The value of Akaike Information Criterion (AIC) is 259.4849. The Residual deviance is 60.604 on 60 degrees of freedom.
According to the values of the above parameters, we found that the number of remaining species has a positive correlation with the area of the island. The larger the area of the island

is, the higher the number of residues on the island is. However, there is a negative correlation between time and the quantity in species remaining. With the passing of the years, the number of remaining species gradually decreases, and the species gradually becomes extinct.

Compared with the previous parameter, in the data of Luke et al. (2013), the value of c is 2.22296, in the new Nonlinear Normal Regression, the value of c is 2.22514, in the Nonlinear Poisson Regression, the value of c is 2.22972, and the z value is in Luke et al. (2013) is less than the value in the new Nonlinear Normal Regression, and the value of the z in the Nonlinear Poisson Regression is higher, than the value in the Nonlinear Normal Regression. This means that when the area of the island increases, the increase in the number of remaining species using the Nonlinear Poisson regression is higher than the result obtained by using the Nonlinear Normal Regression, and the result in Luke et al. (2013). The value of k is similar on all sides, but because of another difference in parameters, it can also be concluded that when the Nonlinear Poisson Regression is fitted, the rate of species extinction is faster, than in the model obtained by using the Nonlinear Normal Regression and Luke et al. (2013) with the Nonlinear Normal Regression Original.

From Table 2 the results of the Parameters Confidence Interval are 2.5 % - 97.5 % respectively s_∞ is from 0.5498 to 1.9824, b is from 0.8264 to 4.9897, z is from 0.7725 to 1.0047, k is from 0.0631 to 0.0976. For the Parameter Confidence Interval, the problem to be solved is to examine how close the estimated parameter value is to the true value of the parameter. The parameter value of the Nonlinear Normal Regression obtained by the fitting is that s_∞ is 1.2192, $w(c/s_\infty)$ is 1.8288, z is 0.8803, k is 0.0780, c is 2.2297, the coefficient of determination is 0.8660 and Aic is 259.4849. Comparing the corresponding parameter confidence interval, we can deduct that all the fitted values are in the parameter confidence interval.

The coefficient of determination is higher than the Nonlinear Normal Regression, and the value of AIC is almost the same as the Nonlinear Normal Regression. The Residual deviance from Nonlinear Poisson Regression is less than the Residual deviance from Nonlinear Normal Regression, so it is more effective.

	2.5 %	97.5 %	Estimate	Pr(>z)	Reject
s_∞ Normal	-1.4691	3.0141	0.9640	0.3949	Yes
s_∞ Poisson	0.5498	1.9824	1.2192	0.0004	No
b Normal	1.0608	3.8145	2.2251	0.0022	No
b Poisson	0.8264	4.9897	1.8288	0.0137	No
z Normal	0.7296	1.0620	0.8719	1.65e-15	No
z Poisson	0.7725	1.0047	0.8803	<2e-16	No
k Normal	0.0545	0.0998	0.0732	5.30e-08	No
k Poisson	0.0631	0.0976	0.0789	<2e-16	No

Table 5-2: Parameter's Confidence Interval and Hypothesis Testing

Parameter's Confidence Interval

According to the Table 2, the Confidence Interval of each parameter in Nonlinear Poisson Regression fitting and the Nonlinear Normal Regression fitting, we need to analyze the difference between the result.

The Parameter Confidence Interval, as the possible range of the value of the parameter, can effectively express the range where the value of the parameter may be. Although the fitted value is a certain value, it may also appear within the range of the Confidence Interval.

Comparing the Confidence Interval of each parameter of the above two different regressions, we cannot tell the quality of these parameters. However, it can be seen from the fitted values obtained that the value of s_∞ is positive, and the Nonlinear Normal Regression model is used to obtain a part of the Confidence Interval where s_∞ is negative and means that this part of the Confidence Interval cannot be expressed effectively. Comparing the value of parameter b with Nonlinear Normal Regression and Nonlinear Poisson Regression, we can see that the Confidence Interval of parameter b is higher in Nonlinear Poisson Regression than in Nonlinear Normal Regression. For parameters z and k, compared with Nonlinear Normal Regression and Nonlinear Poisson Regression, we can see that the Confidence Interval of parameter z and parameter k is similar, but the Confidence Interval of parameter z and parameter k in nonlinear Poisson regression is smaller than the Confidence Interval of parameter z and parameter k in nonlinear Normal Regression. So the Nonlinear Normal Regression is not as effective as the Nonlinear Poisson Regression.

Residual Deviance

From Nonlinear Normal regression, we can get the Residual deviance is 1014 on 60 degrees of freedom. From Nonlinear Poisson Regression, we can get the Residual deviance is 60.604 on 60 degrees of freedom. Less Residual deviance means the regression is more effective. So for Residual Deviance, Nonlinear Poisson Regression is more effective than the Nonlinear Normal Regression.

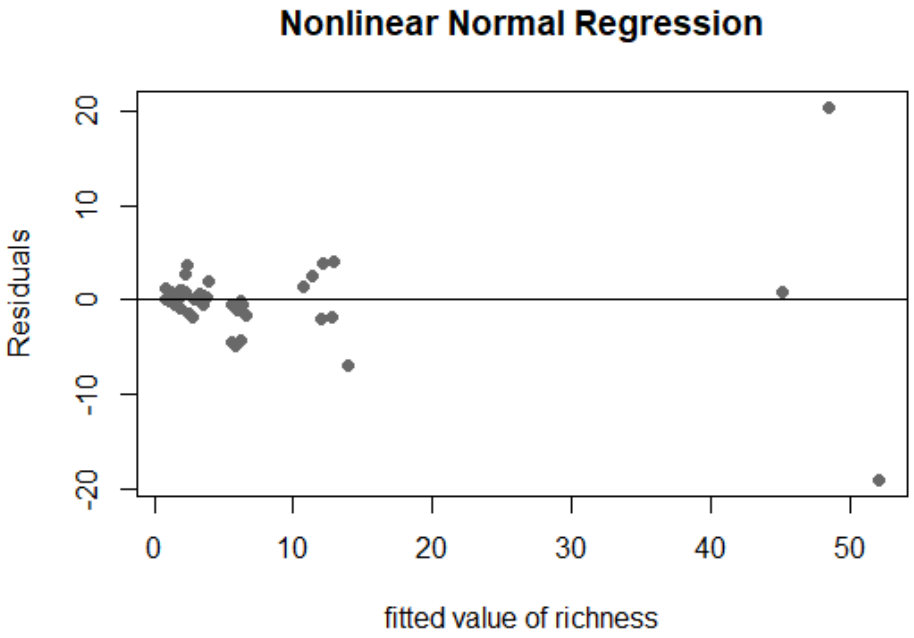

Figure 5-1: Residual vs fitted value Nonlinear Normal regression

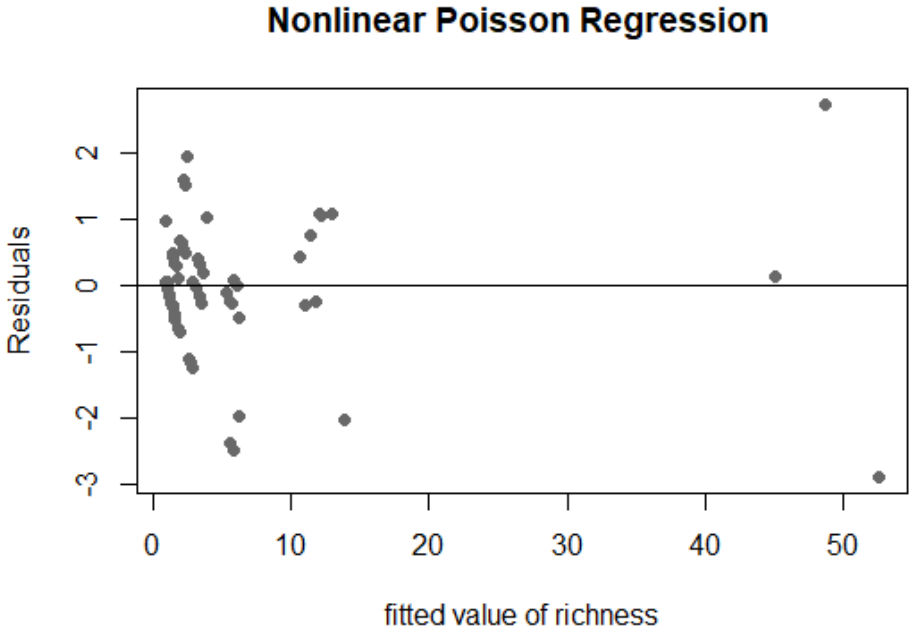

Figure 5-2: Residual vs fitted value Nonlinear Poisson regression

For residual, it is more effective when the point of the data is closer to the line. If it is closer, it is more effective when we compare Nonlinear Normal fitting and Nonlinear Poisson fitting.Concerning Nonlinear Poisson fitting. The points are all at the interval between (-3, 3), most of the points fitted by Nonlinear Normal Regression are between (-5, 5), and there are two points between (-20, 20), and can be seen from the figure. Using Nonlinear Poisson Regression fitting is more effective.

Hypothesis Testing for Parameters

$$\begin{cases} H_0: & \beta_{s_\infty,w,z,k} = 0 \\ H_1: & \beta_{s_\infty,w,z,k} \neq 0 \end{cases}$$

The approximation of the critical value or P-value to help decide is statistic. in this case, it identifies the significant variables through R program language. The P-value is the minimum significance level at which the null hypothesis would be rejected.

Valor P $\leqslant \alpha$ so reject H_0 to the level α, means that the parameter is effective, and can effectively achieve the effect of displaying the fitted data.

Valor P $> \alpha$ so no reject H_0 to the level α, means the parameter is invalid, and the effect of displaying the fitted data cannot be effectively achieved.

Assuming α is 5 %, for the Nonlinear Poisson Regression. According to the Figure 2, the P-value for parameters s_∞ and w and z and k are $\leqslant \alpha$ so reject H_0 to the level α, it means that the parameter s_∞, w, z, k are effective and can effectively achieve the effect of displaying the fitted data.

According to the Table 2, for the Nonlinear Normal Regression. The P-value for parameters w and z and k are $\leqslant \alpha$ so reject H_0 to the level α, means that the parameter w, z, k are effective and can effectively achieve the effect of displaying the fitted data. P-value for s_∞ is $0.395 > \alpha$ so no reject H_0 to the level α, it means that the parameter s_∞ is invalid, and the effect of displaying the fitted data cannot be effectively achieved.

Because Luke et al. (2013) does not contain the data of hypothesis, comparing the Nonlinear Normal Regression model and the Nonlinear Poisson Regression model, the parameter s_∞ and w in the Nonlinear Normal Regression no reject H_0 al nivel α, in Nonlinear Poisson Regression rejects H_0 to the level α, so the conclusion shows that using Nonlinear Poisson Regression is more efficient than Nonlinear Normal Regression.

5.1. Analysis of Remaining Species

Comparison of Remaining Species

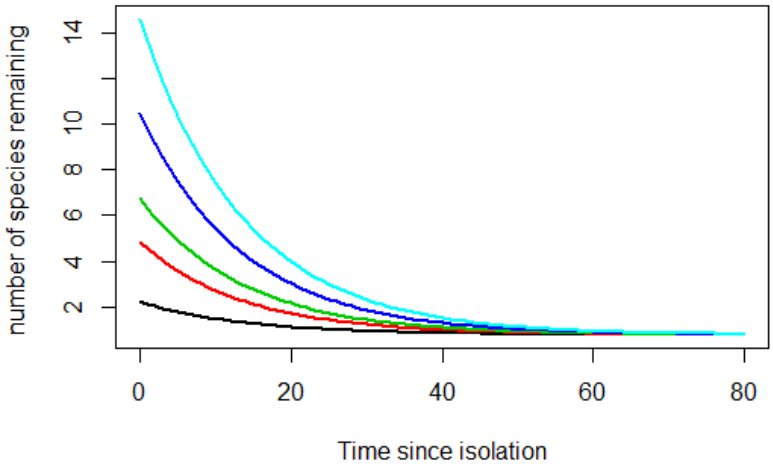

Figure 5-3: Number of species remaining since Time isolation original

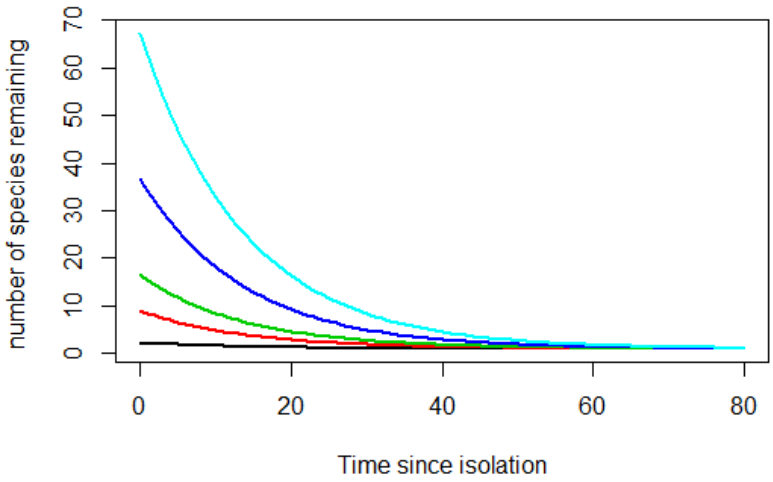

Figure 5-4: Number of species remaining since Time isolation Normal

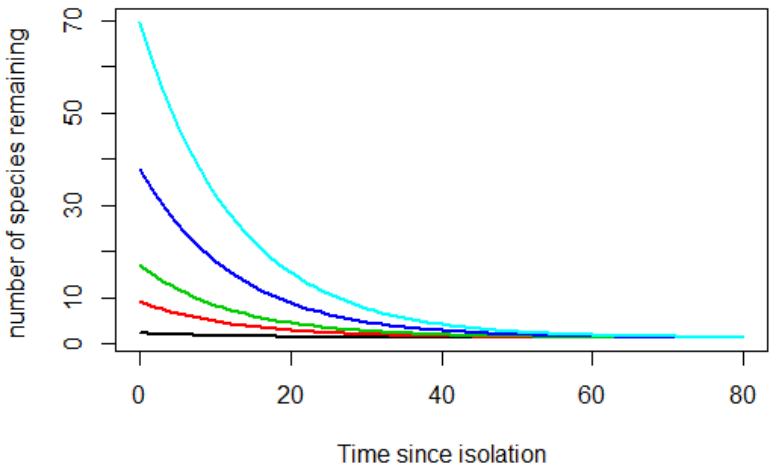

Figure 5-5: Number of species remaining since Time isolation Poisson

With the time passing, in order to get the Curve between the number of remaining species on each island and the time, using Nonlinear Normal Regression and Nonlinear Poisson Regression to fit the data from Luke et al.(2013) to get result respective, also according to the result of the original of the Nonlinear Normal Regression from Luke et al.(2013), set the number of remaining species as the Y axis and time as the X axis. With the passing of the years, the number of remaining species on each island (1 ha, 5 ha, 10 ha, 25 ha, 50 ha) is displayed. Figure 3, 4, 5 show the extinction of the number of remaining species in different islands (1 ha, 5 ha, 10 ha, 25 ha, 50 ha) with time passing, when using the results from Nonlinear Normal Regression and Nonlinear Poisson regression and the original Luke et al.(2013) to estimate. All the figures show that when the time reaches 15 years, remaining species in all the islands are half extinct. When the time reaches 25 years, the remaining species on each island are basically extinct. All the figures also show that the species on larger islands can be extinct faster than in smaller islands with time passing.
The different Curves in the Figure 3, 4, 5 show the islands of different areas (1 ha, 5 ha, 10 ha, 25 ha, 50 ha), the color is from dark blue to light blue. The Curve of light blue is 50 ha, the Curve of dark blue is 1 ha. At the beginning, the number of remaining species in large islands is higher than in small islands, we can see that the larger islands have a higher rate of species extinction than the small islands because of the large number of remaining species. According to the fitting of Nonlinear Normal Regression and Nonlinear Poisson regression and result from the Luke et al.(2013), with time passing, the species from all islands are basically extinct after 40 years. In order to compare the figures according to the result when using Nonlinear Normal regression and Nonlinear Poisson regression and the original from Luke et al.(2013). for these three figures, there have so little difference in each curve (1 ha, 5

ha, 10 ha, 25 ha, 50 ha), The extinction in the original result is faster than in the rest, and is almost similar for Nonlinear Normal Regression and Nonlinear Poisson Regression. Overall, it can be seen that the remaining species have become extinct rapidly in the early stage. The rate of extinction decreases with time passing until the species are basically extinct. The figures also show that the number of species from larger islands is extinct faster than in smaller islands with time passing.

Comparison of the Rate of Remaining Species Extinction

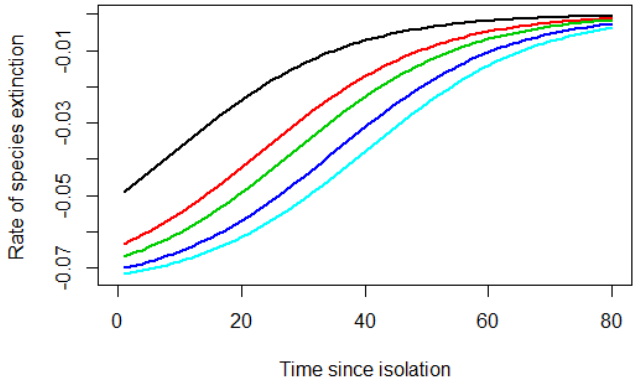

Figure 5-6: Rate of Special Time since isolation base on Original

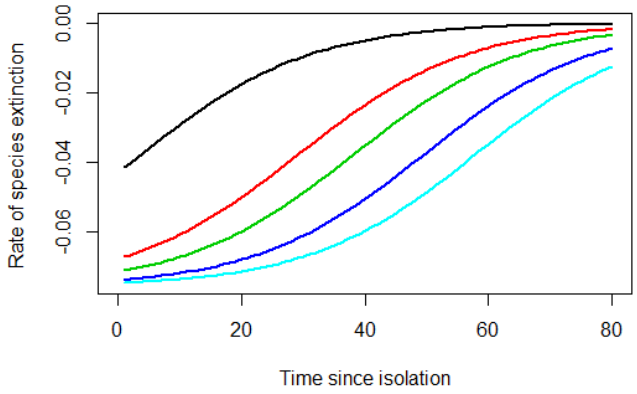

Figure 5-7: Rate of Special Time since isolation base on Normal

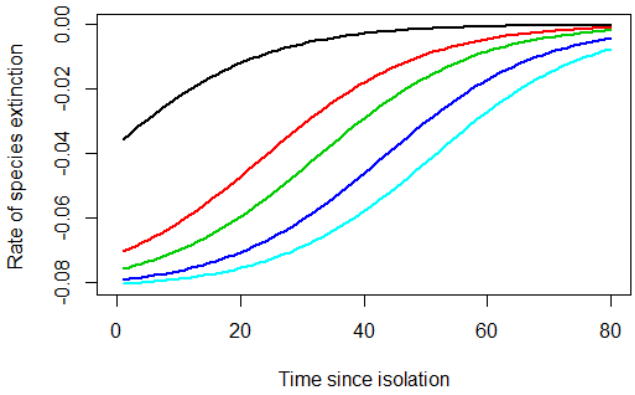

Figure 5-8: Rate of Special Time since isolation base on Poisson

With the time passing, in order to get the Curve between the Rate of Special on each island and the time, using Nonlinear Normal Regression and Nonlinear Poisson Regression to fit the data from Luke et al.(2013) to get result respective, also according to the result of the original from Luke et al. (2013), set the rate of remaining species as the Y axis and time as the X axis. With the time passing, the number of remaining species on each island is displayed.

We analyzed the date from Luke et al. (2013), using Nonlinear Normal regression fitting and Nonlinear Poisson regression fitting. Results showed the Figure 7 and Figure 8. From Figures 7 and 8 show that the results from Nonlinear Normal regression fitting and Nonlinear Poisson regression fitting. From Figures 6 show that the result from the original of Luke et al. (2013), it is also using Nonlinear Normal Regression fitting, but the result is difference, and all the figures show the remaining species extinction rates of different islands are different, and the remaining species extinction rate of large islands is higher than that of small islands.

The rate of extinction of the remaining species in all islands is gradually decreasing with time passing. When the island area is 50 ha, the remaining species extinction rate of the island is higher than other small islands, because the larger island can maintain a better ecological environment. Therefore, it has more number of remaining species, while in the small island, the extinction rate of remaining species tends to be close to 0. The smaller the island cannot have a good ecological environment, so the number of remaining species tends to zero.

The information in the figures show that there is little difference in the fitting from the data of Luke et al. (2013), when using the result from Nonlinear Normal Regression and Nonlinear Poisson Regression and the original of Luke et al. (2013). With the time passing, the rate of remaining species extinction is decreased. The extinction rate decreases with time passing until species are basically extinct. The figures also show that the extinction rate in larger islands is faster than in smaller islands with time passing. But in the end, all the remaining

species will become extinct to zero.

Analysis of Half Time with Mean Value of Island

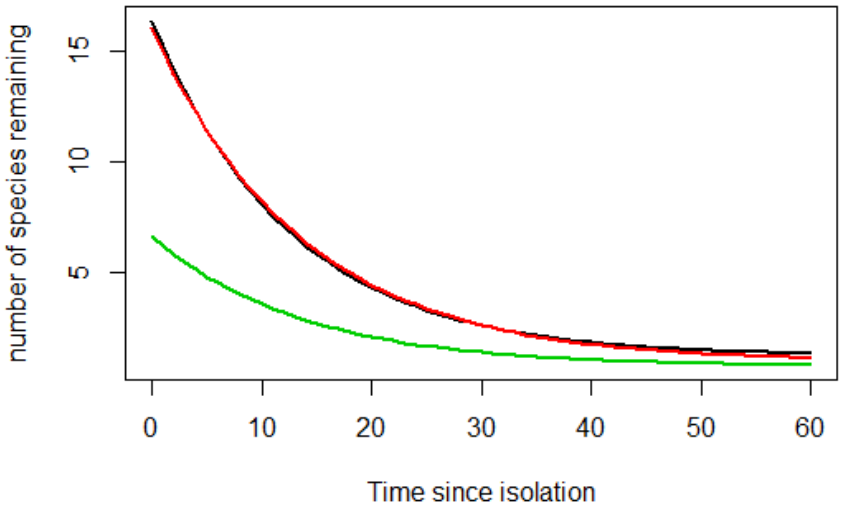

Figure 5-9: Correlation Years and Richness-mean island

The Figure 9 shows that base on the original data from Luke et al. (2013), Nonlinear Normal Regression fitting and Nonlinear Poisson Regression fitting are respectively. Result reveals that, base on the mean of the size of island area, the extinction of the number of the remaining species decrease with the time passing, and species are expected to be completely reduced to zero within 40 years. For the Curve derived from the original article data (Green Curve), the number of remaining species present the original Nonlinear Normal Regression with time passing, and the extinction rate of species is decreasing. When the time passing to the 10 years, the number of remaining species are basically extinct to half. For the mean island, the remaining species are expected to be completely reduced to zero within 25 years.
For the Curve obtained by using Nonlinear Normal Regression (Red Curve), the number of remaining species shows a trend of extinction with time passing, which is slower than the extinction rate of the original data. When the time passing to the 15 years, the number of remaining species are basically extinct to half in the mean island of the date, the number of remaining species is expected to be completely reduced to zero within 40 years. For the Curve obtained using Nonlinear Poisson Regression (Black Curve), the number of remaining species shows a trend of extinction with time passing, and the extinction rate of species is decreased. When the time passing to the 15 years, the number of remaining species are basically extinct

to half, it is expected to be completely reduced to zero within 40 years. Comparing to the three Curves (Green Curve, Red Curve, Black Curve), the Curve red (Normal) and the Curve black (Poisson) are similar, the Curve green(Original) showed that the extinction rate is more higher than the others. But it seems that the extinction rate of original is too high, so the Green Curve is invalid. The Curves obtained by using Nonlinear Poisson Regression and Nonlinear Normal Regression fitting are closer to the natural extinction rate and are more effective.

Analysis of Half Time With Difference Fragment Area (ha)

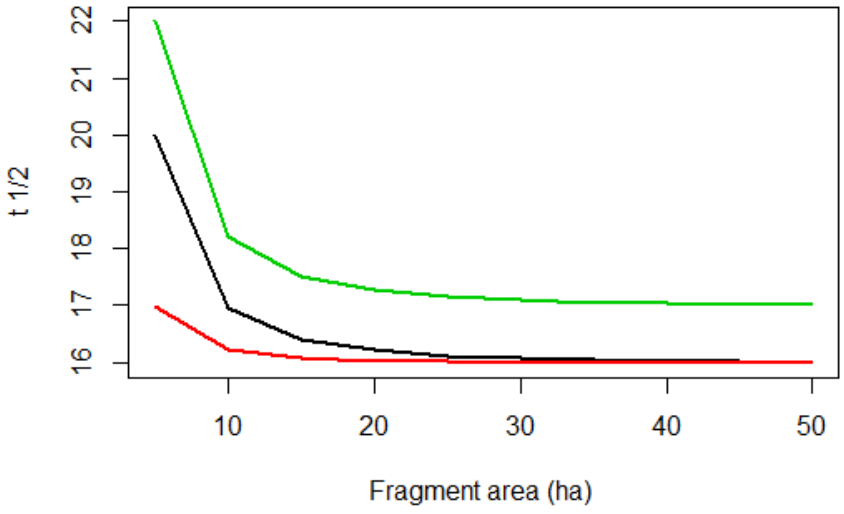

Figure 5-10: Half Time with difference Fragment area (ha)

The Figure 10 shows the curve from the data of the original article Luke et al. (2013), the Nonlinear Normal Regression fitting and the Nonlinear Poisson Regression fitting respectively, the Figure shows that in different island base on the different regression, the time about the remaining species extinct to the half in each size of the island. The green curve is fitting from the Original, the red curve is fitting from the Nonlinear Normal Regression, the black curve is fitting from the Nonlinear Poisson Regression.
Generally, the number of remaining species will be half extinct when the area of the island increases, and the time gradually decreases. The area of the island is divided into several parts, 0 - 5 ha islands, 5 - 10 ha islands, 10 - 40 ha islands, bigger than 40 ha islands, we cannot analyze 0 - 5 ha islands, because the area of the island is too small, so after the island is isolated the number of remaining species on the island is only 2 or 3, there have no

significant to analysis this part. In the island of 5 - 10 ha, we can conclude that the larger the island area is, the number of remaining species will be higher, so the time for species extinction to half will be shorter. For the island of 10 - 40 ha, there is a basic and stable ecological environment, so that the rate of extinction of the remaining is decreasing slowly, but the rate can also be faster than the smaller island. For islands larger than 40 ha, the number of remaining species is large, but the species will be extinct to half faster than the smaller islands.

Comparing the curve of the original data Luke et al. (2013), the curve of Nonlinear Normal Regression and the curve of Nonlinear Poisson Regression, it can be seen that these three lines are similar. But for the time of species extinction to half, the original curve is slower than the Nonlinear Poisson Regression. The Nonlinear Poisson Regression is slower than the Nonlinear Normal Regression. In another hand, for all the curves the larger island is, there will have a faster time of extinct to half, the smaller island is, there will have a slower time of extinct to half.

6 Conclusions and Recommendations

6.1. Conclusions

This type of data (Species Extinction Data Study) is relatively common in biology, including other variables such as area (isolated area or island) and time. The main purpose of this type of data is to estimate the better regression, Y as the number of remaining species and x as the area (isolated area or island) and time etc. If using the Linear Regression, the result for most of this kind of data (Species Extinction Data Study) is not effective, it's better to use the Nonlinear Regression, if the data contains time variable, it is not effective for Normal Regression base on the characteristic, it is effective for the using of Poisson Regression, According to the above the best calculation Statistical method is using the Nonlinear Poisson Regression for this type of data (Species Extinction Data Study).

When using the data from Luke et al. (2013) for regression fitting, according to the result above(Section 5), if using the Nonlinear Poisson Regression, there are less value of $AICs$ and residual and more value of R^2 than using the Nonlinear Normal Regression. The using of the Nonlinear Poisson Regression can be more effective to fit the data, and the using of the Nonlinear Poisson Regression is more reasonable and conducive for data processing.

In most existing biological researches, after studying the better regression between the number of species extinction and their living environment changes, time and other related variables, the biologist found out that with the time passing, the species became extinct, the number of species have the relation positive with the size of area and other kind of variable.

In most of the island fragments, the number of remaining species has decreased with time passing. According to the conclusion of Ferraz et al. (2003) in the Amazon rain-forest, the remaining species in 100 hectares of forest fragments were extinct in less than 15 years. In this experiment from Luke et al. (2013) whether using the Nonlinear Normal Regression or the Nonlinear Poisson Regression, the number of remaining species will be half extinct within 15 - 20 years.

The study also shows that among island fragments of different sizes, bigger island fragments have a larger number of remaining species. With the time passing, the number of remaining species in bigger island fragments, become extinct faster than those in smaller island fragments. After fitting and analyzing with Nonlinear Normal Regression and Nonlinear Poisson Regression respectively, we can conclude that for this type of habitat fragmentation and species invasion data, the using of Nonlinear Poisson regression is more effective.

6.2. Recommendations

As technology advances, human activities in the world continue to increase with time passing, human activities have a more serious impact on the environment. Some human activities have caused area isolation and environmental changes, resulting in a reduction in the number of species. With the increase of these phenomena, more biologists are interested in this type of data, the main research of biostatistics is to analyze this kind of data. This type of data contains time, number of species and area etc, when fitting this type of the data, using the Nonlinear Poisson Regression can be more effective.

As statisticians, it is more important to study how to fit data more effectively, but there also have another important part is that humans need to be more careful about our behavior on the earth, protect the environment, and reduce bad function of environment from human beings. We need to do the best to achieve the balance between human behavior and the environment.

Bibliografía

[1] Palmeirim, Ana Filipa, et al. Small mammal responses to Amazonian forest islands are modulated by their forest dependence. Oecologia 187.1 (2018): 191-204.

[2] Mendes Pontes, Antonio Rossano, et al. Mass extinction and the disappearance of unknown mammal species: scenario and perspectives of a biodiversity hotspot's hotspot. PLoS One 11.5 (2016): e0150887.

[3] Martensen, Alexandre Camargo, et al. Associations of forest cover, fragment area, and connectivity with neotropical understory bird species richness and abundance. Conservation Biology 26.6 (2012): 1100-1111.

[4] Brook, Barry W., Navjot S. Sodhi, and Peter KL Ng. Catastrophic extinctions follow deforestation in Singapore. Nature 424.6947 (2003): 420-423.

[5] Jones, Byron. Statistical Methods for Engineers and Scientists. (1985): 390-390.

[6] Brady, Megan J., et al. Matrix is important for mammals in landscapes with small amounts of native forest habitat. Landscape Ecology 26.5 (2011): 617-628.

[7] Sekercioglu, Cagan H. Conservation ecology: area trumps mobility in fragment bird extinctions. Current biology 17.8 (2007): R283-R286.

[8] Attuquayefio, Daniel K., Erasmus H. Owusu, and Benjamin Y. Ofori. Impact of mining and forest regeneration on small mammal biodiversity in the Western Region of Ghana. Environmental monitoring and assessment 189.5 (2017): 1-9.

[9] Broadbent, Eben N., et al. Forest fragmentation and edge effects from deforestation and selective logging in the Brazilian Amazon. Biological conservation 141.7 (2008): 1745-1757.

[10] Fermin Martin Piera Apuntes sobre Biodiversidad y Conservacion de Insectos: Dilemas, Ficciones y Soluciones? Monografico Los Artropodos y el Hombre (Sociedad Entomologica Aragonesa) 20: 25-55. 1997

[11] Michalski, Fernanda, and Carlos A. Peres. Disturbance-mediated mammal persistence and abundance-area relationships in Amazonian forest fragments. Conservation Biology 21.6 (2007): 1626-1640.

[12] Seber, G.A.F., and C.J. Wild. Nonlinear regression john wiley and sons. New York (1989).

[13] Woinarski, John CZ, Andrew A. Burbidge, and Peter L. Harrison. Ongoing unraveling of a continental fauna: decline and extinction of Australian mammals since European settlement. Proceedings of the National Academy of Sciences 112.15 (2015): 4531-4540.

[14] Deikumah, Justus P., Clive A. McAlpine, and Martine Maron. Biogeographical and taxonomic biases in tropical forest fragmentation research. Conservation Biology 28.6 (2014): 1522-1531.

[15] Keystone Species Hypothesis. University of Washington. Archived from the original on 2011-01-10. Retrieved 2011-02-03.

[16] Liddell, Henry George, Scott, Robert, A Greek-English Lexicon, revisada y aumentada por Jones, Sir Henry Stuart, con la asistencia de McKenzie, Roderick; Oxford, Clarendon Press 1940

[17] Gibson, Luke, et al. Near-complete extinction of native small mammal fauna 25 years after forest fragmentation. Science 341.6153 (2013): 1508-1510.

[18] MacArthur, R. H. y Wilson, E. O. La teoría de la biogeografía de la isla. Princeton, New Jersey: Presión de la universidad de Princeton 1967

[19] Benchimol, Maíra, and Carlos A. Peres. Predicting primate local extinctions within real-world forest fragments: a pan-neotropical analysis. American Journal of Primatology 76.3 (2014): 289-302.

[20] Brady, Megan J., et al. Matrix is important for mammals in landscapes with small amounts of native forest habitat. Landscape Ecology 26.5 (2011): 617-628.

[21] Sharma, Narayan, M. D. Madhusudan, and Anindya Sinha. Local and landscape correlates of primate distribution and persistence in the remnant lowland rainforests of the upper Brahmaputra Valley, Northeastern India. Conservation Biology 28.1 (2014): 95-106.

[22] Sodhi, Navjot S., et al. Deforestation and avian extinction on tropical landbridge islands. Conservation Biology 24.5 (2010): 1290-1298.

[23] Paine, Robert T. A conversation on refining the concept of keystone species. Conservation biology (1995): 962-964.

[24] Chisholm, Ryan A., et al. Species-area relationships and biodiversity loss in fragmented landscapes. Ecology Letters 21.6 (2018): 804-813.

[25] Aguilar, Ramiro, et al. Unprecedented plant species loss after a decade in fragmented subtropical Chaco Serrano forests. PLoS One 13.11 (2018): e0206738.

[26] Huet, Sylvie, et al. Statistical tools for nonlinear regression: a practical guide with S-PLUS and R examples. New York: Springer, 2004.

[27] Philpott, Stacy M., et al. Biodiversity loss in Latin American coffee landscapes: review of the evidence on ants, birds, and trees. Conservation Biology 22.5 (2008): 1093-1105.

[28] W.F.Laurance, M.Goosem, S.G.Laurance, Trends Ecol.Evol.24,659-669 2009

[29] United Nations Environment Programme, What is biodiversity? , World Conservation Monitoring Centre 2010

[30] Laurance, William F., and Diana C. Useche. Environmental synergisms and extinctions of tropical species. Conservation biology 23.6 (2009): 1427-1437.

[31] Newmark, William D., et al. Targeted habitat restoration can reduce extinction rates in fragmented forests. Proceedings of the National Academy of Sciences 114.36 (2017): 9635-9640.

[32] Ferraz, Gonçalo, et al. Rates of species loss from Amazonian forest fragments. Proceedings of the National Academy of Sciences 100.24 (2003): 14069-14073.

[33] Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. rmarkdown: Dynamic Documents for R (version 2.10). https://CRAN.R-project.org/package=rmarkdown. 2021

[34] Draper, Norman R., and Harry Smith. Applied regression analysis. Vol. 326. John Wiley and Sons, 1998.