ISyE 6420

Bayesian Statistics

Chinhsien Tsai

April 25, 2021

Bayesian Linear Regression Model on Predicting the Relationship of Urbanization and Beetles

1. Introduction

Poisson regression models have been widely used with the species abundance. These models take into account the dependent variables with the abundance of species. In this project, the Poisson regression model would be demonstrated by the use of the probability lot. The ecological influence of urbanization is increasing throughout the years. The beetles community has been highly impacted by the human community expansion. In this project, the beetles data collected from New York City by Fusco et al (2017) would be used in the Poisson regression model to investigate the relationship of the beetle species richness(abundance), the type of habitats, and the impervious surface. This model could be used to predict the number beetles species in other city with a specific type of habitats and impervious surface coverage.

2. Data Investigation

In this project, the species abundance, the type of habitats, and the impervious surface coverage are investigated. Based on the raw data in Fusco et al (2017), 2,170 carrion beetles from 8 species in total were collected. There are 13 sites in New York City for beetles collection. Among these sites, three types of habitats, including urban, suburban, and rural were classified. The impervious surface coverage, which indicates how much percentage of natural land coverage is left out of the total land coverage, is also calculated.

Beetles Diversity(Number of Species)	Habitat Type	Impervious Land Coverage
2	1	60
6	1	60
6	1	60
6	1	29
5	1	27
5	2	17
3	2	10
4	2	10
5	2	10
4	2	11
5	2	4
6	2	0
3	3	0
5	3	0
4	3	0

(Raw data from Fusco et al (2017))

3. Model Identification & Results

The Poisson regression is a type of generalized linear model (GLM). It models a positive integer that responses against a linear predictor via a specific link form. It is presented in the form of:

$$\log(\lambda) = \beta_0 + \beta_1 * x_1 + \ldots + \beta_k * x_k$$

The rate lambda is a positive integer. For Poisson regression link is log. The log of lambda links to both positive and negative numbers. In our case:

$$\log\lambda = \beta_1 + \beta_2 * \text{habitat} + \beta_3 * \text{surface}$$

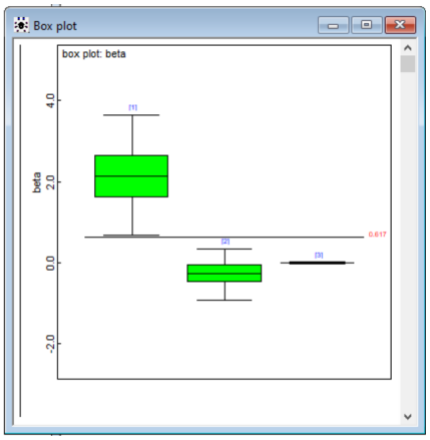
The purpose of this model is to modeling the expectation of the number which would be connect to the lineal combination predictor. Here, parameter lambda is the prediction of the mean number of the beetle species. The coefficient beta[1] is the number of beetle species. Beta[2] is linked to the type of habitat(1=urban, 2=suburban, 3=rural). Beta[3] is linked to the surface coverage.

After getting the result of the log of lambda, the number could be plug in to the Poisson regression to get the predicted number of species at the specific habitat type. For Poisson regression, the deviance is:

$$D = 2\sum_{i=1}^{n}\left(y_i \log\frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i)\right)$$

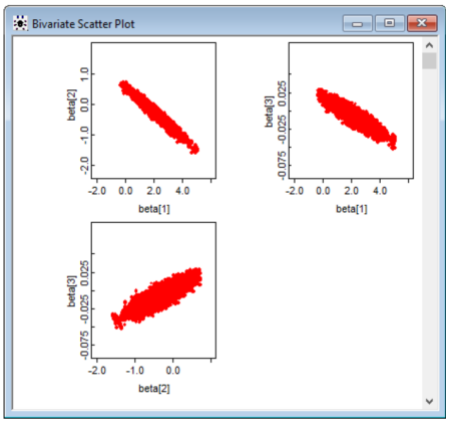
$$\hat{y}_i = \exp\{b_0 + b_1 x_{i1} + \cdots + b_k x_{ik}\}$$

The result shows that the model fits good. In the below image, it shows that beta[1] has the largest positive scale. It also shows that the number of beetle species is the coefficient that plays the most important role in this model.



(Summary of posterior distribution)

In the picture below, it shows that the model fit good. In these three graphs, we could see that the scatter plot are linear relationships. Between beta[1] and beta[2], and beta[1] and beta[3], the relationship are both linear negative relationship. Between beta[2] and beta[3], it is linear positive relationship.



(Scatter plots of coefficients)

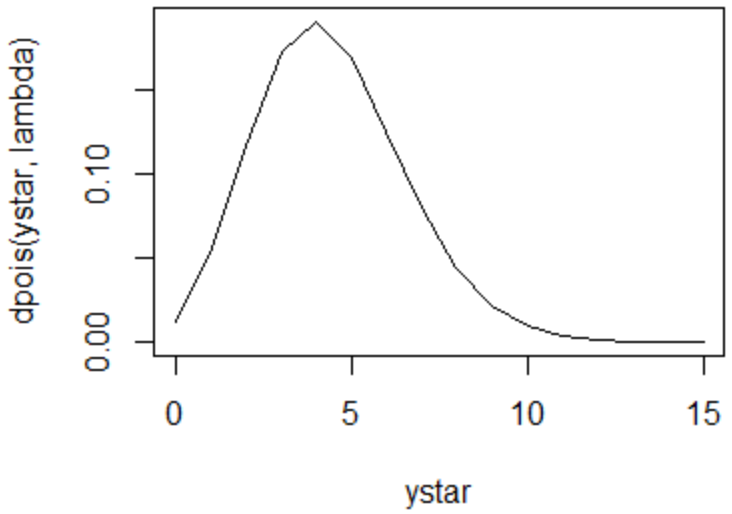
In the picture below, the deviance is 57.92, which is pretty large. However, considering the dataset is small(N=15), it might be the reason why the result is such.

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
beta[1]	2.066	0.7474	0.03753	0.5222	2.107	3.431	1001	100000
beta[2]	-0.2445	0.3083	0.01535	-0.8076	-0.2618	0.377	1001	100000
beta[3]	-0.00553	0.01002	4.475E-4	-0.0249	-0.005703	0.01411	1001	100000
deviance	57.92	2.438	0.06334	55.11	57.3	64.28	1001	100000
lambda[1]	4.593	1.199	0.02245	2.536	4.486	7.214	1001	100000
lambda[2]	4.593	1.199	0.02245	2.536	4.486	7.214	1001	100000
lambda[3]	4.593	1.199	0.02245	2.536	4.486	7.214	1001	100000
lambda[4]	5.397	1.187	0.04889	3.311	5.329	7.937	1001	100000
lambda[5]	5.471	1.263	0.05416	3.272	5.396	8.201	1001	100000
lambda[6]	4.442	0.5612	0.006143	3.41	4.414	5.623	1001	100000
lambda[7]	4.625	0.6371	0.01263	3.476	4.591	5.961	1001	100000
lambda[8]	4.625	0.6371	0.01263	3.476	4.591	5.961	1001	100000
lambda[9]	4.625	0.6371	0.01263	3.476	4.591	5.961	1001	100000
lambda[10]	4.597	0.6159	0.01079	3.477	4.565	5.887	1001	100000
lambda[11]	4.806	0.823	0.02504	3.369	4.766	6.539	1001	100000
lambda[12]	4.94	0.9914	0.03417	3.245	4.876	7.046	1001	100000
lambda[13]	3.918	1.014	0.03461	2.266	3.802	6.186	1001	100000
lambda[14]	3.918	1.014	0.03461	2.266	3.802	6.186	1001	100000
lambda[15]	3.918	1.014	0.03461	2.266	3.802	6.186	1001	100000

Knowing beta[1], beta[2], and beta[3], we could predict number of beetle species in certain type of habit with certain surface coverage. For instance, the number of species in urban habitat and 60 surface coverage would be:

$$\mathbf{lambda} = \mathbf{exp}\,(\mathbf{2.066} - \mathbf{0.2445} - \mathbf{0.00553} * \mathbf{60})$$



4. Conclusion

In the project, the result has shown that the Poisson regression could be perform on the dataset of the beetle species and its relationship with the level of urbanization. However, there are still a few limitations on my project.

First of all, the data set I used is small. I would try a larger dataset, preferably with $N > 50$ in the future study, in order to get a more reliable conclusion to see if the model is really fit or not.

Secondly, the model I used may only be for this sample of data. I might need to use this model to test other type of dataset with different variables to see if it is stable and could predict precisely for different kind of data.

5. Reference

Fusco, N. A., Zhao, A., & Munshi-South, J. (2017). Urban forests sustain diverse carrion beetle assemblages in the New York City metropolitan area. PeerJ, 5, e3088. https://doi.org/10.7717/peerj.3088