YYÜ TAR BİL DERG
(YYU J AGR SCI)
2018, 28(3): 266-270

Geliş tarihi (Received): 24.02.2018
Kabul tarihi (Accepted): 17.09.2018
doi: 10.29133/yyutbd.398359

Araştırma Makalesi/*Research Article (Original Paper)*

Modeling Relative Risk Assesment for Infected Plants by Eriophyoid Mites (Acari, Prostigmata) Using Poisson Log Linear Regression Model

Abdullah YEŞİLOVA[1*], Evsel DENİZHAN[2], Sultan ÇOBANOĞLU[3]

[1]Van Yüzüncü Yıl University, Agricultural Faculty, Department of Biometry,Van, Turkey
[2]Van Yüzüncü Yıl University, Agricultural Faculty, Department of Plant Protection,Van, Turkey
[3]Ankara University, Agricultural Faculty, Department of Plant Protection, Ankara, Turkey
*e-mail: yesilova@yyu.edu.tr

Abstract: In Poisson regression, the dependent variable of the mite population relative risk assessment can be estimated as based on countable data. Due to disappearing data and numerous uncountable values of the mite population it became difficult to evaluate the risk factor by linear regression methods. The assessment of variable features of mites depending on conditions is very suitable for Poisson regression modeling system. In the this study, the occurrence of rare events such as the occurrence ratio of infected plants was defined by eriophyid mites on two wheat varieties and four different localities. The study was constructed by two way possibility table depending on plant varieties (*Triticum aestivum* L. and *Secale cereale* L. (Poaceae) with four locations (Muradiye, Ahlat, Erciş, Doğu Beyazıt and Iğdır). The reference parameters were *Triticum aestivum* for varieties, and Muradiye for location, respectively. The risk assessment of infected plants for *Secale cereale* is 1.245 times higher as compared to *Triticum aestivum* and this difference was found statistically significant ($p<0.05$). The risk of infected plants for Iğdır location is 1.101 times higher as compared to Muradiye location ($p>0.05$). In the Poisson log-linear regression, the dependent variable is a risk ratio or a relative risk can be estimated as well as countable data. Thus, Poisson log-linear regression model is a very effective method for analysis of two-way contingency table. Two-way contingency table is created considering the eriophyid mite infection ratio depending on location and varieties, respectively.

Keywords: Acari, Eriophyoid, Mites, Poisson log-linear regression, Relative risk

Poisson Log Doğrusal Regresyon Kullanılarak Eriophyoid Akar (Acari, Prostigmata) Tarafından Enfekte Edilen Bitkiler için Relatif Riskin Modellenmesi

Öz: Poisson regresyonunda, akar populasyonu için relatif risk tahminini esas alan bağımlı değişken sayıma dayalı olarak elde edilebilir. Bu tip sayıma dalayı olarak elde edilen akar populasyonu için bilinen doğrusal regresyon yöntemi kullanılarak risk tahmini oldukça zordur. Böyle durumlarda, Poisson log doğrusal regresyonu oldukça uygundur. Bu çalışmada, enfekte edilen bitkilerin olma oranları iki farklı buğday çeşidi ve dört farklı lokasyon tarafından tanımlandı. Çalışmanın veri seti; iki farklı buğday çeşidi (*Triticum aestivum* L. and *Secale cereale* L. (Poaceae) ve dört farklı lokasyon (Muradiye, Ahlat, Erciş, Doğu Beyazıt and Iğdır) kullanılarak iki yönlü tablo halinde verilmiştir. Referans parametresi olarak çeşit için *Triticum aestivum* ve lokasyon için ise Muradiye alınmıştır. *Secale cereale*'de bitkilerin enfekte olma riski *Triticum aestivum*'dan 1.245 kat daha yüksek olduğu saptanmıştır ($p<0.05$). Iğdır lokasyonundaki bitkilerin enfekte olma riski Muradiye lokasyonundan 1.101 kat daha yüksek olduğu saptanmıştır ($p>0.05$). Poisson log doğrusal regresyon modelinde bağımlı değişken sayımla elde edilen risk oranı ya da relatif risk biçiminde tanımlanabilir. Bu nedenle Poisson log doğrusal regresyonu bu tip iki yönlü tabloya dönüştürülebilen veriler için oldukça etklili bir yöntemdir. Burada, iki yönlü tablonun oluşumu için çeşitler ve lokasyonlar kullanılmıştır.

Anahtar kelimeler: Akar, Eriophyoid, Poisson log doğrusal regresyon, Relatif risk

Introduction

The Eriophyoidea (Acari: Prostigmata) is one of the largest group of plant-feeding arthropods. Some of these could have a scientific and economic importance of agricultural crops and some of them are on weed control.

Only 130 eriophyoid species have been discovered in Turkey until 2015 (Denizhan et al. 2015). These eriophyoid species have been recorded and described by several authors and published as catalogue (Denizhan et al. 2015). The knowledge of the eriophyoid fauna that still remains undiscovered completely is limited in many parts of Turkey. . As we know some Eriophyoid mites are diverse mostly on host plants and some of them are host plant specific. Due to the geographical position and botanic history of Turkey; it is important, to find out potential range of host plants for these organisms, their damage level and distribution. They cause abnormalities and gall on plants. Besides they are vectors of important virus diseases. They live on the grooves of the upper leaf surface, often causing leaf curling. They are vectors of the Wheat Streak Mosaic Virus (WSMV) and the High Plains Viruses (HPV) (Denizhan et al. 2013).

Eriophyoids are very small organisms. Their distribution does not show normal population characteristics and it will be more accurate to detect them by Poisson distribution method in determining whether they are risky in agricultural areas or plants. Therefore, evaluation of the mite risk assessment by linear regression is not suitable for modeling mite populations based on count data (Eye and Mun 2013). Linear regression requires dependent variables to be in a normal distribution (Yeşilova and Denizhan 2016). It is very difficult to determine changes in population in mite populations due to variations such as temperature, plant variety, etc. It is very difficult to determine with normal population models and may not reflect actual situation (Yeşilova and Denizhan 2016). It is important to use Poisson regression for assessment of the Risk assessment values for temperature based on different sampling localities and the plants varieties. In the Poisson regression, the dependent variable shows a Poisson distribution (Agresti 1997; Yeşilova et al. 2010). The advantage of the distribution is that, it allows quantitative rather than qualitative analysis of the values observed in the condition-dependent variables in the mite population, and is based on countable data.

Besides, the evaluation of the relative risk (RR) ratio of the mite population could be estimated by the Poisson regression (Eye and Mun, 2013). Dependent variable is very appropriate for modeling the events such as infected plants data set.

Our goal is to identify the Eriophyid species on Wheat plants in Van Region, Turkey. Thus, the aim of this study was to determine the relative risk ratio of the mites based on different wheat plants (*Triticum* and *Secales*) with different locations (Ağrı, Iğdır, Muradiye and Doğu Beyazıt) by applying Poisson log-linear regression model.

Material and Methods

Sampling was conducted in Eastern and South-Eastern parts of Turkey in different locations (Ağrı, Iğdır, Muradiye and Doğu Beyazıt) and varieties (Triticum and Secales) from 2013 to 2014 as a part of a faunistic survey. The wheat samples were collected from agricultural areas and weeds around Van Lake Region between 2013 and2014. The sampling, classic mounting and identification techniques were used. The numbers of Eriohyoids were counted from wheat samples belonging to four different localities for two weeks interval in Van,Turkey in 2013-2014. The plants were examined under a stereomicroscope for the presence of eriophyoid mites. Mite specimens were collected, mounted on slides using a standard protocol (De Lillo et al. 2010) and then identified (Keifer 1969; Amrine et al. 2003)

Poisson log linear Regression Analysis

Poisson log linear model

Mean of Poisson distribution is μ. We supposed that the population has an "n" number of plants and a risk that a plant will be infected is "p." Then the average count of plants that have been infected is $\mu = n * p$.

We write proportion for infected plant as $p = \frac{\mu}{n}$.

y_{ij} is the observed count in the (i, j) cell

n_{ij} is the population in the (i,j) cell for i = 1,...,r and j = 1,..., c. Thus, $p_{ij} = y_{ij} / n_{ij}$

		Factor B			
		1	2	3	c
Factor A	1	y_{11} n_{11}	y_{12} n_{12}	y_{13} n_{13}	y_{1c} n_{1c}
	2	y_{21} n_{21}	y_{22} n_{22}	y_{23} n_{23}	y_{2c} n_{2c}
	:	:	:	:	:
	r	y_{r1} n_{r1}	y_{r2} n_{r2}	y_{r3} n_{r3}	y_{r3} n_{rc}

$p_{ij} = y_{ij} / n_{ij}$

$$\log\,[\,p_{ij}] = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k$$

Then

$$p_{ij} = e^{\beta_0 + \beta_1 X_1 + ... + \beta_k X_k}$$

and

$$\hat{p}_{ij} = e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + ... + \hat{\beta}_k X_k}$$

This formulation helps to clarify the meaning of the maximum likelihood coefficients: $e^{\hat{\beta}_i}$ gives the change in the risk for Y when there is a unit change in the predictor X_i, i = 1,..,k

An adjusted RR between two categories of the variable, (X_1) is

$$\hat{RR}_{X_1=1 \text{ vs } X_1=0} = \frac{\hat{\text{Risk}}(X_1 = 1, X_2, ..., X_k)}{\hat{\text{Risk}}(X_1 = 0, X_2, ..., X_k)} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 X_2 + ... \hat{\beta}_r X_k}}{e^{\hat{\beta}_0 + \hat{\beta}_2 X_2 + ... \hat{\beta}_r X_k}} = e^{\hat{\beta}_1}$$

Analysis of Maximum Likelihood

The Wald Chi-Square Test is to test $H_0 : \beta_i = 0$ for $i = 1,...,k$ vs H_A : Not H_0.

$$W = \frac{\hat{\beta}_i^{\,2}}{\hat{SE}(\hat{\beta}_i)^2}, \quad i = 1,...,r; \quad j = 1,...,c$$

In our study, we suppose the following data set are classified by two factors: varieties (with two levels) and location (with five levels) period.

Poisson log-linear model is,

$$\text{Log}\,(p_{ij}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_3 Z_3$$

where X_1 for Ahlat region,, X_2 for Doğu Beyazıt region, X_3 for Erciş region, and X_4 for Iğdır region; Z3 = 1 for Secale cereale. We constructed two way possibility table between plant varieties (*Triticum aestivum* and *Secale cereale*) and location (Muradiye, Ahlat Erciş, Doğu Beyazıt and Iğdır) in this study. The reference parameters were as follow: *Triticum aestivum* for varieties, and Muradiye for location.

Results

The required statistical analyzes were performed by using SAS 9.1.4 statistical software program. The number of infected plants depending on the plant varieties and the localities were considered as dependent variable in the model. The varieties (*Triticum aestivum* and *Secale cereale*) and locations (Muradiye, Ahlat Erciş, Doğu

Beyazıt and Iğdır) were included in the model as dependent variables. Poisson log-linear regression was applied to data set using proc GENMOD procedure in SAS. Maximum likelihood parameter estimates for Poisson log linear regression was given Table1. In Table 1, the effects of location levels were found statistically insignificant on the number of infected plants ($p>0.05$). The effects of varieties were found statistically significant on the number of infected plants ($p<0.01$).

Table 1. Parameter estimates for Poisson log linear regression

Analysis Of Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Confidence**	**95% Limits**	**Chi-Square**	**Pr > Chi-Sq**
Intercept		1	-0.322	0.1025	-0.523	-0.123	9.88	0.002
Location	Ahlat	1	0.068	0.132	-0.190	0.327	0.27	0.603
Location	Doğu Beyazıt	1	0.062	0.132	-0.198	0.321	0.22	0.642
Location	Erciş	1	-0.002	0.131	-0.259	0.255	0.00	0.987
Location	Iğdır	1	-0.097	0.136	-0.364	0.170	0.50	0.479
Location	Muradiye	0	Reference parameter				.	.
Varieties	Secale cereale	1	-0.219	0.086	-0.38	-0.049	6.41	0.011
Varieties	Triticum aestivum	0	Reference parameter				.	.

Relative risk statistics for each location and variety level according to reference parameters were calculated using table 1. Thus, RR statistic values were,

$$RR_{Erciş\ vs\ Muradiye} = e^{-0.0020} = 0.998\ and\ \frac{1}{0.998} = 1.002$$

$$RR_{Iğdır\ vs\ Muradiye} = e^{-0.0966} = 0.907\ and\ \frac{1}{0.907} = 1.101$$

$$RR_{D.Beyazıt\ vs\ Muradiye} = e^{0.068} = 1.063\ and\ \frac{1}{1.063} = 0.940$$

$$RR_{Ahlat\ vs\ Muradiye} = e^{0.0685} = 1.071 and\ \frac{1}{1.071} = 0.934$$

$$RR_{Sc\ vs\ Ta} = e^{-0.219} = 0.803\ and\ \frac{1}{0.803} = 1.245$$

In the Poisson regression, the dependent variable is a risk ratio or a relative risk can be estimated as well as data set based on count data. Dependent variable and the occurrence of rare events assessment such as infected plants ratio are very suitable for modeling Poisson risk evaluation. Sampling were conducted at the coastal band of Van Lake basin in 2013-2014. We constructed two way contingency measurement table between varieties (*Triticum aestivum* and *Secale cereale*) and locations (Muradiye, Ahlat Erciş, Doğu Beyazıt and Iğdır) in this study. In regression coefficient particular level of each independent variable is compared to reference category. There is a difference between the linear and nonlinear tests. In this study, the reference parameters taken were as follow: *Triticum aestivum* for varieties, and Muradiye for location. The risk of infected plants for *Secale cereale* is 1.245 times higher as compared to *Triticum aestivum* and this difference was found statistically significant ($p<0.05$) (Table 1). The risk of infected plants for Ahlat location is 0.934 times less then as compared to Muradiye location ($p>0.05$), the risk of infected plants for Ahlat and Doğu Beyazıt is 0.940 times less than as compared to Muradiye location ($p>0.05$), the risk of infected plants for Erciş location is 1.002 times higher as compared to Muradiye location ($p>0.05$), and the risk of infected plants for Iğdır location is 1.101 times higher as compared to Muradiye location (p>0.05). It was determined that all the comparisons of locations were statistically insignificant according to reference category (Table 1).

Discussion and Conclusion

The determination of the biodiversity of the plant parasitic Eriophyoids on wheat and its economic importance is rich and substantial. The role of the eriophyoid species on wheat and its economical importance should be investigated in Turkey (Denizhan et al. 2013). Especially the host-parasite relationship should be evaluated by concerning the monophagi of the plant parasitic eriophyoid species. There should be different evaluation and Modeling determination system of the risk factors specially to control the important invasive mites of plant species (Keifer 1969; Amrine et al. 2003). Therefore, a large survey on the eriophyoid mites in Van lake basin with particular regard to the species on the wheat and their evaluation systems should be developed. It is necessary to develop new statistical methods or suitable assessment models. Thus there will be more informative methods for the assessment of the invasive Eriophyoid population and its damage level.

Poisson log-linear regression is based on generalized linear models (GLM). Without requiring any assumption on the original structure of a data set, Poisson regression is defined in the exponential form in GLM. In GLM, regression analysis is applied by using a link function that is suitable for the distribution of dependent variable. Log link function is used for Poisson regression in GLM (SAS 2017).

Chi-sqaure analysis are used to analyze the two-way contingency table, but it is insufficient for a larger contingency table. Thus, Poisson log-linear regression can be used to determine more than two-way contingency table (SAS 2017). However, in the Poisson log-linear regression, the dependent variable is a risk ratio or a relative risk can be estimated as well as count data (Agresit 1997; Eye and Mon 2013). Dependent variable is quite suitable for modeling the occurrence of rare events such as cancer and infected plant data set.

References

Agresti A (1997). Categorical Data Analysis. John and Wiley & Sons, Incorporation, New Jersey, Canada.

Amrine J W Jr (2003). Catalog of the Eriophyoidea. A working catalog of the Eriophyoidea of the world. Ver. 1.0. <http://in- sects.tamu.edu/research/collection/hallan/acari/eriophyidae> (accessed 1 July 2013).

De Lillo E, C Craemer, J W Amrine, Jr, Nuzzaci G (2010). Recommended procedures and techniques for morphological studies of Eriophyoidea (Acari: Prostigmata). Exp. Appl. Acarol. 51: 283–307.

Denizhan E, Szydło W, Skoracka A (2013). Eriophyoid studies in Turkey: Review and perspectives. Biol. Lett. 50: 45–54.

Denizhan E, Monfreda R, De lillo E, Çobanoğlu S (2015). Eriophyoid mite fauna (Acari: Trombidiformes: Eriophyoidea) of Turkey: new species, new distribution reports and an updated catalogue. Zootaxa, 3991(1), 1-63.

Eye A V, Mun E Y (2013). Log-Linear Modeling. John and Wiley & Sons, Inc., Hoboken, New Jersey, Canada.

SAS (2017). SAS/Stat Software Hangen and Enhanced, SAS Institute Incorporation, USA.

Yeşilova A, Kaydan B, Kaya Y (2010). Modeling insect-egg data with excess zeros using zero-inflated regression models. Hacettepe Journal of Mathematics and Statistics, 39(2), 273-282.

Yeşilova A, Denizhan E (2016). Modeling mite counts Using Poisson and negative binomial regressions. 25(11), 5062-5066. Fresenius Environmental Bulletin.