International Journal of Engineering & Technology, 7 (3.28) (2018) 68-71

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Stepwise Logistic Regression Analysis: An application toward Poultry Farm Data in Johor

Siti Afiqah Muhamad Jamil[1]*, Mohd Asrul Affendi Abdullah[1], Kek Sie Long[1], Nur Fazilla Mohd Jupri[2], Mustafa Mamat[3]

[1]*Faculty of Applied Science and Technology, Universiti Tun Hussein Onn Malaysia, Bandar Universiti, 84600 Pagoh, Johor, Malaysia*
[2]*Faculty of Computer and Mathematical Science, Universiti Teknologi MARA, Wisma KUB, 15050, Kota Bharu, Kelantan, Malaysia*
[3]*Faculty of Computing and Applied Mathematics, Universiti Sultan Zainal Abidin, Besut Campus, Terengganu, Malaysia*
**Corresponding author E-mail: hw150021@siswa.uthm.edu.my*

Abstract

The aims of this study are to fit a logistic regression model towards the fly problem in a farm and to identify the variables that are associated with the fly problem in a poultry farm. By using SPSS software, this study used 'FORWARD STEPWISE' and 'BACKWARD STEPWISE' methods to perform the analysis. Compared to linear regression analysis, logistic regression does not require rigorous assumptions to be met. This study used Likelihood Ratio test, Omnibus test and Hosmer and Lemeshow test to validate and to test the fit of poultry farm data. Akaike Information Criterion (AIC) is calculated to observe the difference between the methods of stepwise used by SPSS software in this study. As a result, logistic regression is fit towards poultry farm data by a stepwise procedure. BACKWARD STEPWISE seems to be more suitable for conducting the stepwise method of analysis. Besides, variables that influence the problem of fly in a poultry are the wasps, distance and number of flies.

Keywords: *Stepwise Logistic Regression Analysis; SPSS software, Akaike Information Criterion (AIC).*

1. Introduction

The term of poultry means a variety types of birds that is raised for food, fibre or others. Poultry plays an important role in Malaysia as it is part of Malaysian diet. People consume meat, eggs or broth in their daily life. Poultry industry in Malaysia has become very important as the product is imported to other country especially to Singapore and Thailand [7].

Poultry meat is the main protein product for all ethnic group in Malaysia and it is dominant product in all food service outlets. The price of beef and goat or even pork is more expensive compared to the price of chicken. Since it has low price, the production of chicken is more compared to others.

Based on [8], poultry sub-sector in Malaysia contributed about RM4.0 million which involved the production meat and eggs. Besides, in any livestock industry, it is the large proportion of production cost. The animal feeds that are having raw ingredient are not produced in Malaysia but it is imported such as the corn gluten meal, bone meal, soybean meal and others. Continuously, the livestock feed production projected or increased from 1996 to 2008. For another 5 years, it is predicted to be increased at an average of 2.8%.

Poultry operation need to be controlled as the waste material produced causes a greatly increase in the fly problems [3]. Besides, if the production of poultry farm did not manage well, external parasite might slow the growth and reduce the production of egg and lower the validity [5]. There are three types of fly which are the house fly, black garbage fly and little house fly. They are from the original breeding source of the location of egg, larvae and pupae produced. Larvae is an immature form of an insect as it is between stages of egg and pupae. The temperature and the species of fly are the main factors of fly problem. Besides, the house fly is known to have between 8 to 20days of life span and within 25 miles away. It can disperse under an average summer conditions [1]. The black garbage fly about 4 to 5 mile away from their source of breeding. One of the programmes named Integrated Fly Control Management programmed (PKLB) by the government has been managed to control the flies' problem in poultry farm either by using biological, chemical or the technical control [6].

The logistic regression method of analysis has been used by the researcher as the results involved a response that is in binary form of data. It is known as the simplest method compared to probit method of analysis [4]. Compared to probit model, the interpretation by using logistic regression model is easier. Besides, there is no assumptions that specifically need to be met as the dependent variable is in dichotomous values.

In addition, in order to choose the best model of analysis, this study performed stepwise logistic regression analysis. There are two methods of stepwise method can use either Wald statistic, likelihood ratio or conditional algorithm for the removal of variables, it is relevant to increase the accuracy of the final model [2].

Consequently, this study has two aims in order to complete the procedure of analysis which are to fit a logistic regression model towards the fly problem in a farm. This study used SPSS software to complete the analysis.

2. Methodology

The study used secondary data collected from Department of Veterinary Service Johor (DVSJ), which comprised the data of a poultry farm data.

Table 1: Description of Data

Variables	Description
Flies Problem (Categorical: 0-No Problem, 1-Problem)	The problem of flies in a poultry farm
Coop system (Categorical: 0-Open system, 1-Closed system)	The system of a poultry farm either open house or closed house.
Coop management (Discrete)	Number of times farmer clean the poultry water, food container and roof infrastructure in a week
Farm Type (Discrete)	Type of poultry farm: (1-Poultry meat, 2-Poultry egg, 3-Poultry duck, 4-Poultry village, 5-Quails)
Beetle (Discrete)	Number of times farmer releasing beetle in a week
Wasps (Discrete)	Number of times farmer releasing wasps in a month
Foliar spray (Discrete)	Number of times farmer spraying fungi (probiotic) by using yeast/ actinomycetes in a week
Larvicide spray (Discrete)	Number of times farmer spraying larvicide that contain cyromazine/ diflubenzuron in a week
Residue spray (Discrete)	Number of times farmer spraying larvicide that contain cyhalotrin/ alpha-cypermethrin in a month
Drainage (Discrete)	Number of times farmer clean all water drainage in a week
Wreckage (Discrete)	Number of times farmer manage poultry wreckage in a week
Distance (Continuous)	Distance between chicken coop in meter
Tiles (Continuous)	Distance of tiles from the poultry manure in feet

The list of variables used for the analysis is shown. It describes the management of a poultry towards the problem of fly in poultry farm. In order to perform a complete procedure of analysis, the framework of analysis is shown in Fig. 1.

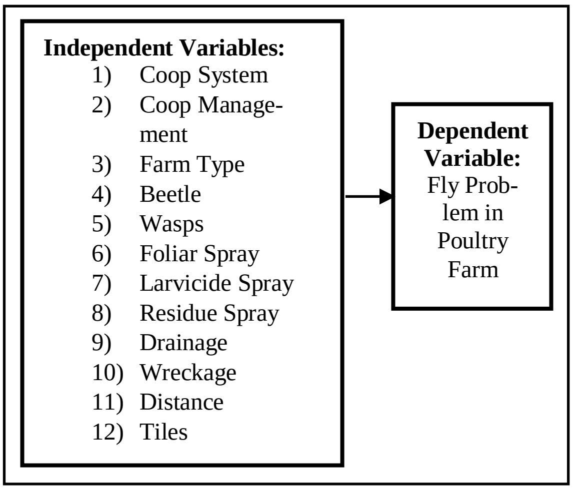

Fig. 1: Framework of data

In order to analyse which of the variables associates with the flies' problem in a poultry farm, this study performed a stepwise logistic regression analysis. The dependent variable is in binary form of data. By starting with description of data, this study performed the 'FORWARD STEPWISE' followed with 'BACKWARD STEPWISE' method of analysis.

3. Results and Discussion

Based on Table 2, the descriptive statistics shows overall frequency of each variables. The farm has closed the house about 114 times compared to open the house with 86 times. Frequently, the farm has been cleaned 3 times every week. Besides, this study manages mostly poultry meat in the farm compared to other farm type. Beetle mostly released 3 times, while wasps about 2 times. Foliar spray and larvicide sprayed mostly 4 times a week, while residue has been sprayed mostly 3 times per month. Drainage has been managed mostly 4 times, while wreckage about 5 times per week. Minimum distance is about 5 meters and the maximum is 25 meters.

Table 2: Descriptive Statistics

Var	0	1	2	3	4	5	6	7	8
1)	114	86							
2)		13	11	40	38	39	33	26	
3)		63	35	13	21	18			
4)	22	18	50	65	45				
5)	23	28	59	56	34				
6)		1	45	62	77	15			
7)		1	40	65	76	18			
8)	5	7	27	49	43	43	21	4	1
9)		24	32	49	70	24			
10)				2	40	62	48	48	
11)	Min: 5		Max: 25		Mean: 15.08		Variance: 21.949		
12)	Min: 0		Max: 18.40		Mean: 8.1762		Variance: 29.876		
13)	Min: 498		Max: 9865		Mean: 3830.75		Variance: 4482646.63		

In order to find the significant variables that influenced the problem of flies' in the farm, this study has performed a stepwise logistic which is also a binary logistic by using stepwise regression analysis. Both methods start with no variables included in the model.

Table 3 shows the classification of accuracy of the flies' problem, which is about 62%. Classification met high accuracy when the value is greater than 80%.

Table 3: Table of Classification

Observed		Predicted		
		flyproblem		Percentage Correct
		0	1	
flyproblem	0	124	0	100.0
	1	76	0	00.0
Overall Percentage				62.0

Table 4 shows the beginning model before entering the variables. Without predictors (variables), the *P*-value = 0.001 < α =0.005, the model is significant.

Table 4: Beginning Block

Variables in Equation					
B	S.E.	Wald	df	Sig.	Exp(B)
-0.490	0.146	11.293	1	0.001	0.613

3.1. FORWARD stepwise logistic regression

The accuracy of the flies' problem increased from 62% to 82.5% that indicates a high classification of accuracy, which can be seen in Table 5.

Table 5: Forward Classification Table

Step 4		Predicted		
		flyproblem		Percentage Correct
		0	1	
flyproblem	0	109	15	87.9
	1	20	56	73.7
Overall Percentage				82.5

Table 6 shows the fourth step of forward logistic regression after the variables are included in this study. Since the *P*-value = 0.039 < α =0.005 with Chi-square value is 10.080, this indicates the model has poor fit.

Table 6: Omnibus Tests

		Chi-Square	df	Sig.
Step 4	Step	10.080	4	0.039
	Block	106.175	10	0.000
	Model	106.175	10	0.000

The step of forward logistic can be seen in Table 7. The log likelihood values are decreasing while the Cox and Snell R-Squared and Nagelkerke R-Squared increased in value. The small value of likelihood indicates the data fit logistic distribution.

Table 7: Summary of Model

Step	-2 Log likelihood	Cox and Snell R Square	Nagelkerke R Square
1	196.128	0.294	0.399
2	182.632	0.340	0.462
3	169.531	0.382	0.519
4	**159.451**	**0.412**	**0.560**

From Table 8, since the *P*-value = 0.238 > α =0.005 indicates that the model has quite good fit after performing the forward logistic regression analysis.

Table 8: Hosmer Lemeshow

Step	Chi-Square	df	Sig.
1	37.853	8	0.000
2	13.308	8	0.102
3	6.225	8	0.622
4	**10.408**	**8**	**0.238**

The variables that enter in each step of forward logistic regression are:

a. Variable(s) entered on step 1: numflies.
b. Variable(s) entered on step 2: distance.
c. Variable(s) entered on step 3: wasps.
d. Variable(s) entered on step 4: farmtype.

The final model of forward logistic regression is shown in Table 9.

Table 9: Forward Stepwise Model

Variables in Equation					
B	S.E.	Wald	df	Sig.	Exp(B)
-2.103	1.096	3.679	1	0.055	0.122
0.001	0.000	44.362	1	0.000	1.001
-0.158	0.049	10.474	1	0.001	0.854
		14.910	4	0.005	
		9.176	4	0.057	

3.2. BACKWARD stepwise logistic regression

The accuracy of the flies' problem increased from 62% to 83.5% that indicates a high classification of accuracy which can be seen in Table 10.

Table 10: Backward Classification Table

Step 4		Predicted		
		flyproblem		Percentage Correct
		0	1	
flyproblem	0	109	15	87.9
	1	18	58	76.3
Overall Percentage				83.5

Table 11 shows the fourth step of forward logistic regression after the variables are included in this study. Negative value of Chi-square indicates the value has decreased from the previous. Since the *P*-value = 0.134 > α =0.005 with Chi-square value is 124.093, this indicates the model has a good fit towards the data of poultry farm.

Table 11: Onmibus Tests

		Chi-Square	df	Sig.
Step 4	Step	-7.042	4	0.134
	Block	124.093	20	0.000
	Model	124.093	20	0.000

There are eight steps of backward stepwise logistic that can be seen in Table 12. The log likelihood values are increasing while the Cox and Snell R-Squared and Nagelkerke R-Squared decreased in value. The value of log likelihood need to be small to get a good fit of the models and poultry data.

Table 12: Summary of Model

Step	-2 Log Likelihood	Cox and Snell R Square	Nagelkerke R Square
1	114.215	0.531	0.722
2	114.988	0.529	0.720
3	115.309	0.528	0.719
4	117.629	0.523	0.711
5	126.134	0.502	0.683
6	126.988	0.500	0.680
7	134.490	0.481	0.654
8	**141.532**	**0.462**	**0.629**

From Table 13, since the *P*-value = 0.659 > α =0.005 indicates that the model has quite good fit after performing the backward stepwise logistic regression analysis.

Table 13: Hosmer Lemeshow

Step	Chi-Square	df	Sig.
1	8.174	8	0.417
2	4.754	8	0.784
3	6.394	8	0.603
4	6.576	8	0.583
5	4.625	8	0.797
6	4.992	8	0.758
7	16.769	8	0.033
8	**5.898**	**8**	**0.659**

All the variables are entered in the first step. Then, the variables are remove one by one in each step.

a. Variable(s) removed on step 2: wreckage.
b. Variable(s) removed on step 3: coopsys.
c. Variable(s) removed on step 4: larvicidespray.
d. Variable(s) removed on step 5: residuespray
e. Variable(s) removed on step 6: tiles
f. Variable(s) removed on step 7: drainage
g. Variable(s) removed on step 8: foliarspray

The final model of backward logistic regression is shown in Table 14.

Table 14: Model Summary

	Model	1	2
Stepwise Regression Analysis	Method	Forward	Backward
	-2 Log likelihood	159.451	141.532
	Cox and Snell R Square	0.412	0.462
	Accuracy	82.5%	83.5%
	AIC	1339.516	1333.338

As a summary, backward method shows a better result compared to forward method of stepwise regression analysis. So that, the best model goes to the backward stepwise regression analysis.

4. Conclusion

Binary logistic regression analysis could be applied when the data of dependent variable appear to be in binary form. Continuously, the assumptions of regression analysis such as normal distribution could not be met in linear or multiple regression analysis.
Therefore, the forward and backward regression analysis, showed a good fit towards the method of stepwise analysis. This can be seen based on the Hosmer Lemeshow test of analysis. By forward and backward method, since the *P*-value = 0.238 > α =0.005, and *P*-value = 0.659 > α =0.005 respectively, indicates that both method of logistic model has quite good fit towards the poultry farm data. Thus, based on the final model, the variables that are significant towards the fly problem of a poultry farm are wasps, distance and number of flies whereas, backward method is the best model compared to forward method of analysis. Overall, it was observed that flies problem exists in poultry farm cause of the wasps, distance and number of flies are variables that significant towards the fly problem.

Acknowledgement

We would like to thank the Ministry of Higher Education, Universiti Tun Hussien Onn Malaysia for supporting this research project and Department of Veterinary Service Johor (DVSJ) that help us to get the poultry farm data.

References

[1] Arroyo, H.S (1998). "House fly, Musca domestica Linnaeus (Insecta: Diptera: Muscidae)" Journal of Food and Agriculture Science.

[2] Whittingham, M.J., Stephens, P.A., Bradbury, R.B., & Freckleton, R.P. (2006). "Why do we still use stepwise modelling in ecology and behaviour?" Journal of Animal Ecology, 5(75), 1182-1189.

[3] Loftin, K.M., Hopkins, J.D., & Corder, R. (2014). "Biology and Control of Flies in Poultry Facilities" Agriculture and Natural Resources.

[4] Wuensch, K.L. (2014) "Binary logistic regression with SPSS" 2015.

[5] Zsak, L., Strother, K.O., & Kisary, J. (2008). "Partial genome sequence analysis of parvoviruses associated with enteric disease in poultry" Avian Pathology, 4(37), 435-441.

[6] Kumara, T.K., Disney, R.H.L., Hassan, A.A., Flores, M., Hwa, T.S., Mohamed, Z. & Bhupinder, S. (2012). "Occurrence of oriental flies associated with indoor and outdoor human remains in the tropical climate of north Malaysia" Journal of Vector Ecology, 1(37), 62-68.

[7] Jalaludin, S. (1992). "The development of poultry in Malaysia" Proceedings of the 6th Anim. Sci. Congr. of the Asian-Australian association of Anim. Prod. Soc, 4, 14-21.

[8] Loh, T.C. (2004). "Livestock production and the feed industry in Malaysia" FAO Animal Production and Health Proceedings (FAO). FAO.