Using of Logistic Regression in Animal Science

Article *in* Journal of Applied Sciences · October 2005

DOI: 10.3923/jas.2005.1753.1756 · Source: DOAJ

4 authors, including:

Ecevit Eyduran
Igdir Üniversitesi
170 PUBLICATIONS **690** CITATIONS

Bahattin Çak
Yuzuncu Yil University
18 PUBLICATIONS **38** CITATIONS

Emre Alarslan
Sheep Research Institute, Bandirma,Turkey
7 PUBLICATIONS **13** CITATIONS

Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 5 (10): 1753-1756, 2005
ISSN 1812-5654

Using of Logistic Regression in Animal Science

[1]E. Eyduran, [1]T. Ozdemir,[2]B. Çak and [1]E. Alarslan
[1]Biometry Genetics Unit, Department of Animal Science, Faculty of Agriculture,
Department of Animal Science, Faculty of Veterinary, University of Yüzüncü Yil, 65080 Van, Turkey

Abstract: This study was carried out to examine the effects of environmental factors on different growth with Chi-square, G-test and logistic regression analysis after body weights of these growth periods were categorized as binary. Besides, logistic regression was also based on concordant statistics except Chi-square and G-tests as model fit criteria. With respect to three fit criteria, the relationships among categorized birth weight with categorized body weights in 45th, 60th and 75th days were significant ($p<0.01$). Moreover, the relationship between periods of the lambs born in early March of 2001 year by using logistic regression. The relationships between categorized body weights and categorized birth weight and/or environmental factors were analyzed sex and only categorized body weight in 75th day was significant ($p<0.05$). It could be said that birth weight, one of environmental factors, improved model fit more than the other factors when considered all model fit criteria in logistic regression.As a result, it can be suggested that in addition to Chi-square and G-tests used for providing relationship between two traits, logistic regression in terms of obtaining from different information will be an alternative analysis in place of variance analysis.

Key words: Chi-square test, G-test, logistic regression, animal science

INTRODUCTION

Logistic Regression (LR) has been studied recently in different scientific fields such as medicine, economy and agriculture. LR is a multivariate technique where dependent variable is only categorical (generally binary) and independent variables, mixtures of categorical and continuous variables. LR can be prepared due to some reasons: there are no assumptions on distributions of independent variables, residual terms and on heterogeneity of variances[1].

Considered on animal science, effects of environmental factors such as sex, birth type, birth weight, genotypes and dam age on different growth periods are important[2].

This study aimed to examine the effects of sex, birth type and birth weight on body weights at different periods by using LR instead of variance analysis and to assess model fit criteria such as Chi-square test G test and concordant test using for LR.

MATERIALS AND METHODS

Animal materials of this study were composed of 45 Hamdani lambs raised in Van province of Turkey. Data of sex (26 male and 19 female), birth type (32 single and 13 twin) a nd body weights at different periods (birth weight, body weights at 45th, 60th and 75th days) of the lambs born in early March of 2001 year were recorded.

Body weights at these periods categorized as binary (low and high) are presented in Table 1. Male sex and single birth type were determined as reference categorizes of independent variables in this study. Data obtained from this study were analyzed with procedure freq and logistic of SAS program[3].

Logistic regression is a part of generalized linear model. Dependent variable consists of data coded as binary (0 and 1) (Table 2) while independent variables compose of grouped or categorized data (0, 1, 2) and continuous data. Category coded as 1 in dependent variable, called as reference category, is risk factor of interest[4].

The LR model can explain the effects of independents variables on binary dependent variable.

$$\text{Logit}\{P(Y=1|X)\} = \beta_0 + X'\beta \qquad (1)$$

Where, β_0 is the intercept parameter and β is the vector of slope parameters. Occurrence probability of

Corresponding Author: Ecevit Eyduran, Biometry Genetics Unit, Department of Animal Science, Faculty of Agriculture, Yüzüncü Yil University 65080, Van, Turkey, Tel: +904322251024/2648

Table 1: Categorized ranges of body weights at different periods

Traits	Low body weights (less than) (kg)	High body weights (more than) (kg)
Birth weights	2.28	2.29
Body weights at 30th day	4.67	4.68
Body weights at 45th day	5.85	5.86
Body weights at 60th day	7.33	7.34
Body weights at 75th day	9.38	9.39

Table 2: Cross tabulation of variables

	Explanatory Variable (Sex)	
Dependent variable	Male (X=1)	Female X=0
High weight (Y=1) of body weight at 75th	a = 15	b = 5
Low weight (Y=0) of body weight at 75th	c = 11	d = 14

reference category in dependent variable, denoted by P (Y=1), can be written as follows:

$$P(Y=1 \mid X_i) = \frac{e^{\beta_0+\beta_1X_1+...+\beta_pX_p}}{1+e^{\beta_0+\beta_1X_1+...+\beta_pX_p}} \quad (2)$$

P(Y=0), other category of dependent variable, can be also written as:

$$P(Y=0 \mid X_i) = 1 - P(Y=1 \mid X_i) = \frac{1}{1+e^{\beta_0+\beta_1X_1+...+\beta_pX_p}} \quad (3)$$

Odds ratio (O), the ratio of Eq. 2 and 3, are given in Eq. 4 :

$$\frac{P}{1-P} = e^{\beta_0+\beta_1X_1+...+\beta_pX_p} \quad (4)$$

After transformed to log-scale, Eq. 4 can be rewritten as follow:

$$\text{Logit}(P) = \log \frac{P}{1-P} = \beta_0 + \beta_1X_1 + ... + \beta_pX_p \quad (5)$$

It can be said, therefore, the Multiple Regression Analysis (MRA) is analogous to logistic regression as seen in Eq. 5[1,5].

The association between odds ratio (O) and the probability of being P (Y=1) can be expressed as:

$$P = \frac{O}{1+O} \quad \text{or} \quad O = \frac{P}{1-P} \quad (6)$$

Where, P is the probability of risk occurrence, denoted normally by P (Y=1) and O is odds ratio[4].
For example:

$$\text{Odds ratio} = \frac{\dfrac{P(\text{Body weight at 75th day} \mid \text{Male})}{1\text{-}P(\text{Body weight at 75th day} \mid \text{Male})}}{\dfrac{P(\text{Body weight at 75th day} \mid \text{Female})}{1\text{-}P(\text{Body weight at 75th day} \mid \text{Female})}}$$

$$\text{Odds ratio} = \frac{\text{Odds (Body weight at 75th day} \mid \text{Male)}}{\text{Odds (Body weight at 75th day} \mid \text{Female)}}$$

$$\text{Odds ratio} = \frac{a/c}{b/d} = \frac{15/11}{5/14}$$

$$\text{Odds ratio} = \frac{P}{1\text{-}P} = e^{1.3398} = \frac{a.d}{b.c} = \frac{15*14}{5*11} = 3.818$$

The odds of being high weight in terms of body weight at 75th day for male were 3.818 times higher than that for female.

Model fit criteria: Likelihood ratio chi-square (called also as G-test) (Eq. 7) and chi-square statistics (Eq.8), generally known as goodness of fit criteria, are used for assessing whether model fits or not in logistic regression[4-6]. Expressions of both statistics, also used for many multivariate statistics techniques, can be briefly written as follows:

$$G = 2 \quad f.\ln \frac{f}{f_i} \quad (7)$$

$$\chi^2 = \frac{(f-f_i)^2}{f_i} \quad (8)$$

Where, f observed frequency and f_1; expected frequency, respectively.

Concordant: The criterion is a measure of association between observed responses and the predicted probabilities[1].

RESULTS AND DISCUSSION

It was observed in Table 3 that associations of birth type only with body weight at 30th day and 60th day (p<0.01) as well as association of sex only with body weight at 30th day were significant (p<0.05). As examining Table 4, associations of birth weight with body weight at 30th day, 60th day and 75th day, respectively, were much significant (p<0.01).

As examining Table 5, odds of being high weight at 75th day for male was 3.818 times higher than that for female (p<0.05).

Odds of being high weight at 75th day for being single birth type was 8.038 times higher than that for being twin birth type (p<0.05).

Table 3: Associations of sex and birth type with body weights at different periods

Trait pairs	G-Test	Probability of G-test	Chi-square test	Probability of Chi-square test
Birth Weight -Sex	1.673	0.1959	1.666	0.1968
Birth Weight -Birth Type	1.630	0.2017	1.625	0.2025
Body weight at 30th day -Sex	0.073	0.7871	0.073	0.7871
Body weight at 30th day-Birth Type	3.545	0.0597	3.380	0.0660
Body weight at 45th day -Sex	0.276	0.5997	0.275	0.6001
Body weight at 45th day -Birth Type	7.791	0.0053**	7.188	0.0073**
Body weight at 60th day -Sex	3.500	0.0614	3.411	0.0648
Body weight at 60th day -Birth Type	10.379	0.0013**	8.935	0.0028**
Body weight at 75th day -Sex	4.500	0.0339*	4.377	0.0364*
Body weight at 75th day - Birth Type	3.545	0.0597	3.380	0.0660

*:p<0.05, **:p<0.01

Table 4: Associations of birth weight with body weights at different periods

Trait pairs	G-Test	Probability of G-test	Chi-square test	Probability of Chi-square test
Birth Weight - Body weight at 45th day	12.758	0.0004**	12.068	0.0005**
Birth Weight.- Body weight at 45th day	9.085	0.0026**	8.668	0.0032**
Birth Weight - Body weight at 45th day	10.824	0.0010**	10.286	0.0013**

*:p<0.05, **: p<0.01

Table 5: Individual effects of sex birth type and birth weight on body weights at different periods.

Variables	Degrees of freedom	Estimation of parameter	Standard error	Wald statistics	Probability of wald statistics	Odds ratio	Concordant values (%)
Body weight at 75th day							
Sex (Male)	1	1.3398	0.6550	4.1841	0.0408	3.818*	42.00
Birth type (Single)	1	1.3291	0.7475	3.1613	0.0754	3.778	34.00
Birth weight	1	3.4564	1.2749	7.3505	0.0067	31.703**	76.60
Body weight at 60th days							
Sex (Male)	1	1.1838	0.6528	3.2880	0.0698	3.267	39.7
Birth type (Single)	1	2.7362	1.1001	6.1862	0.0129	15.429*	43.7
Birth weight	1	3.5735	1.2994	7.5632	0.0060	35.641**	76.9
Body weight at 45th days							
Sex (Male)	1	0.3185	0.6081	0.2743	0.6005	1.375	28.4
Birth type (Single)	1	2.0842	0.8488	6.0295	0.0141	8.038*	41.5
Birth weight	1	3.0468	1.1961	6.4892	0.0109	21.049*	76.4

*: p<0.05, **: p<0.01

Odds of being high weight at 60th day for being single birth type was 15.429 times higher than that for being twin birth type (p<0.05). It could be said that lambs of being single birth type had higher weight with respect to others. Examining in concordant test, known also as model fit criteria except G and Chi-Square tests in LR, additions of factors such as sex a nd birth type as independent variables on model fit had low-moderate level (28.4-43.7). On others hand, addition of birth weight in model on model fit had also high level with respect to sex birth type (about 77).

It is known that effects of sex, birth type and birth weight on body weights at different periods are important[2]. This study was to interpret odds ratio in cross tabulation and to examine in model fit criteria such as G, Chi-Square and Concordant in LR.

Birth type on body weights at 45th day and 60th day in terms of G test and Chi-square test were effective. Odds of being high weight at 45th and 60th days for single birth type were 8.038 and 15.429 times than those of twin birth type, respectively (p<0.05) (Table 4). Addition of the factor on model fit was moderate level.

Effect of sex on only body weights at 75th day was significant (p<0.05) (Table 3 and 5). Odds of being high weight at the periods for male lambs were 3.818 times than that for female lambs.

With respect to results of G test and Chi-square test, relationships between birth weight with respectively body weight at 30th day, 60th day and 75th day were much significant (p<0.01).

Contrary to other factors, birth weight factor improved model fit at all periods (Table 3 and 4). When a unit increasing in Birth weight realized, odds of being high weight at those periods were much high and significant (Table 5).

CONCLUSIONS

As a result, it can be suggested that calculation of Chi-Square test with G test in cross tabulations should be needed. With those tests, both using as model fit criteria in LR and proving relationship between two traits are possible. Besides, researchers may obtain much information by LR instead of variance analysis.

REFERENCES

1. Sharma, S., 1996. Applied Multivariate Techniques. John Wiley and Sons. Inc., pp: 317-341.
2. Düzgüneş, O., A. Eliçin and N. Akman, 1996. Animal Breeding. University of Ankara, Faculty of Agriculture Publishing. 3rd Edn., Ankara, pp: 19-27.
3. SAS, 1998. SAS Institute Inc. Cary. NC. USA.
4. Allison, P.D., 1999. Logistic Regression Using the SAS System: Theory and Application. Cary, NC: SAS Institute Inc., pp: 11-40.
5. Eyduran, E., 2005. Logistic regression analysis. Doctorate seminar. University of Yuzuncu Yil, Institute of Natural Applied Science, Van, pp: 3.
6. Sokal, R.R and F.J. Rohlf, 1981. Biometry. The Principles and Practice Of Statistics in Biological Research. 2nd Edn., W.H. Freeman and Company. New York, pp: 731-738.