Aquaculture Research, 2007, **38**, 635–643 doi: 10.1111/j.1365-2109.2007.01707.x

Logit models for evaluating spawning performance of channel catfish, *Ictalurus punctatus* (Rafinesque)

Herbert E Quintero[1], **Asheber Abebe**[2] **& Donald Allen Davis**[1]

[1]Department of Fisheries and Allied Aquacultures, Auburn University, Auburn, AL, USA
[2]Department of Mathematics and Statistics, Auburn University, Auburn, AL, USA

Correspondence: D Allen Davis, Department of Fisheries and Allied Aquacultures, Auburn University, 203 Swingle Hall, Auburn, AL 36849-5419. E-mail: davisda@auburn.edu

Abstract

Broodstock evaluations are often measured by variables such as spawning success, fecundity, fertilization and hatching rates, usually expressed as percentage values. Outcomes are generally analysed as continuous random variables, assuming that they follow a normal distribution. Ordinary linear regression models (e.g. analysis of variance) as well as χ^2 analysis are typically applied. However, these models may not be the most appropriate as a number of test criteria may not be met. For example, spawning success outcomes are inherently discrete and non-negative data and hence their distribution is not likely to be normal. As these models may not be the most appropriate, a case study using logit analysis as an alternative method for the evaluation of this type of data is presented by considering the response as binary data (spawned versus did not spawn). An exact version of logit analysis was performed due to the sparseness of the data. The results demonstrate that appropriate statistical models provide better insight into the cause–effect relationships that exist between control variables and the dependent variable (likelihood of spawning in this case). As would be expected, each strain of fish responded somewhat differently to the test variables. Changing the protein level of the diet from 32% to 42% or increasing the feeding frequency from three to six times per week either did not influence spawning or negatively affected spawning respectively. Additionally, older fish performed better than younger fish and the early spawning period was better than the later spawning period, regardless of strain. These responses, however, were only detected using logit analysis, which is a more sensitive test and would thus be recommended for this type of data.

Keywords: logistic regression, catfish broodstock, spawning analysis

Introduction

Aquaculture development has been feasible due to a combination of factors, such as the application of controlled reproduction, which often rely on induced spawning or hypophysation, environmental manipulations and cryopreservation (Bromage 1995; Donaldson 1996). Multiple interrelated factors are involved in reproduction success including the type and doses of hormones used in hypophysation, the size and age of broodstock, condition factor, the nutritional status of broodfish and environmental conditions. Knowledge of these factors is useful in improving fry production. Spawning success is one of the primary outcomes of this process and it has typically been evaluated using two approaches. The first approach is to consider it as a continuous variable and to further assume that it is normally distributed. In this case, ordinary linear regression models are applied, such as analysis of variance (ANOVA), following some transformation such as an arcsine transformation (Lambert 1998; Celada, Antolin, Carral, Pérez & Sáez-Royuela 2006; Peck & Holste 2006; Rodríguez-González, García-Ulloa, Hernández-Llamas & Villareal 2006). The second approach is to use the Pearson χ^2 test for association in a cross-tabulation where the independent variable is expressed as the number of successes in the spawning process (Bondari, Ware, Mullinix Jr & Joyce 1985). However, the use of ordinary linear regression techniques for a dichotomous-dependent variable could be considered inappropriate from a theoretical standpoint (Allison 1999) and interpretation of the results is not straightforward due to the transformation applied on the response variable. Pearson's χ^2 analysis tests for association between variables but is unable to give estimates of parameters (Stokes, Davis & Koch 2000).

Spawning success can be considered a *discrete random variable* instead of a *continuous random variable* as the possible measured or assigned values consist of a discrete set of categories i.e. organisms that spawned successfully versus those that did not spawn. The analysis of continuous outcome data using linear regression models requires five assumptions: (1) the response is a linear function of the independent variables plus a random disturbance term or random error, (2) the independent variables and random error are independent, (3) variance of the random error is the same for all observations, (4) the random error for one observation is uncorrelated with the random error for any other observation and (5) the errors are normally distributed with a mean of 0 (Allison 1999). Many statistical models make a distinction between response (or dependent) variables and explanatory (or independent) variables. In the classical linear models for regression and ANOVA, one of the variables is treated as the response variable to be explained by the other explanatory variables (Friendly 2000). In this case, the explanatory variables may be quantitative or categorical but the response variable must be quantitative. As spawning success would be a response variable, and it is categorical in nature, the normality assumption imposed on the model residuals is violated. This invalidates the direct application of ordinary linear regression. Additionally, transformations (e.g. arcsine) often do not yield normally distributed data and may make the interpretation of regression coefficients difficult because they are not estimated on the original scale (Byers, Allore, Gill & Peduzzi 2003).

For a dichotomous response variable, such as spawning success coded as 0 for no spawn and 1 for spawn, it is convenient to construct a model relating a function of the probability of successful spawning, say p, to a linear combination of the explanatory variables. A model such as this is commonly known as a generalized linear model (GLM). Our model is the logit model, which is in the family of GLMs. Some of the advantages of this method are that the estimated probability of spawning is guaranteed to take values between 0 and 1, regardless of the values taken by the independent variables (Zelterman 1999, 2002), and that inference for the odds and odds ratios are derived directly from the estimated logit model. The odds of an event are the ratio of probability that it will occur to the probability that it will not occur. As the odds ratio is the ratio of two odds, it provides a simple way of describing the dependence of the factors on each other (Allison 1999; Zelterman 2002).

The probabilistic approach for spawning success can also be achieved from a biological standpoint. Spawning is a result of a maturation process (Heino, Dieckmann & Godø 2002) and typically maturation has been evaluated as a deterministic process. This means that sexual maturity is fully determined by the age and size of an individual. However, maturation is a complex process that is also influenced by other factors such as resource availability and body reserves (Bernardo 1993). As a result, there will be some variability that cannot be explained by the size of maturing individuals at a given age. This highlights the probabilistic nature of the maturation process. Similarly, a spawning response would not occur with certainty once a hormone injection is given. There are other factors, such as water quality and handling stress, which, if maintained properly, lead to ovulation and successful stripping. On the other hand, stress can cause egg retention in the body, causing over-ripening, which indicates failure in the process. In this paper, we investigate the factors that significantly affect successful spawning within a probabilistic framework.

Logit analysis has previously been utilized in ecology, fisheries and eco-toxicology studies, among others. Kellogg, Ligotino and Jinks (1984) developed a logistic regression model to predict thermal mortality in striped bass, *Morone saxatilis* (Walbaum). Munger, Wilde and Follis (1994) determined the relationship between size and age at sexual maturation in flathead catfish, *Pylodictis olivaris* (Rafinesque). DeMartini and Lau (1999) used this analysis to classify fish as either immature or mature, based on body length, ovary weight and oocyte volume for two snappers *Etelis carbunculus* (Cuvier) and *Pristipomoides sieboldii* (Bleeker). Recently, Keating and Cherry (2004) have reviewed logistic regression use and interpretation for wildlife habitat selection studies. Finally, McHugh and Budy (2004) developed spawning habitat suitability models for Chinook salmon, *Oncorhyunchus tshawytscha* (Walbaum), in Idaho. As demonstrated with other species, logit analysis could provide a valuable tool for understanding catfish spawning success. Consequently, this case study was conducted to evaluate the applicability of logit analysis as compared with ANOVA, ANCOVA and χ^2 analyses for the interpretation of spawning data.

Table 1 Number of female channel catfish by strain per treatment

Treatment	Protein (%)	Feed frequency	Strain		Total
			High	Low	
1	42	6	44	55	99
2	42	3	55	57	112
3	32	6	44	56	100
4	32	3	44	59	103
Total			187	227	414

Methods

Data were obtained from a project performed to improve reproductive efficiency by producing hybrids from female channel catfish, *Ictalurus punctatus* (Rafinesque), with male blue catfish, *I. furcatus* (Lesueur), through nutrient manipulations. This study evaluated the influence/interaction of dietary protein level and feeding rate on egg production of two separate strains of catfish (each one with three ages). Two diets containing 42% and 32% protein level (ARKAT Feeds, Dumas, AK, USA) and two feed frequencies (6 and 3 times/week), were used, giving rise to four treatment combinations (TC): TC 1–42% protein level, six feedings per week; TC 2–42% protein level, three feedings per week; TC 3–32% protein level, six feedings per week; and TC 4–32% protein level, three feedings per week. Females were spawned during three periods (early, middle and late season). Evaluation of the dietary protein level and feeding rate treatments were performed through spawning success, total number of eggs and fertility (number of eggs/kg female).

Experimental fish

A total of 414 female channel catfish were maintained at the North Auburn Experimental Station, Auburn University. All broodstock were stocked in 16 ponds, using four ponds per treatment. The females were divided into two strains on the basis of prior spawning behaviour: high-spawning strain (strain 1) and low-spawning strain (strain 2); based on this characteristic, they were assigned proportionally in a randomized manner to each pond (Table 1). Female body weights ranged from 0.4 to 3.4 kg for strain 1 and 0.6 to 3.1 kg for strain 2, with a combined mean weight of 1.74 kg. The organisms were stocked in February 2004, in 0.04 ha ponds at a density of $\sim 1130\,\mathrm{kg\,ha^{-1}}$, giving an acclimation period of approximately 1 month. The trial period was 70–90 days depending on the spawning period. Feeding was carried out during the warmest part of the day between 15:00 and 17:00 hours, at a rate of 1.7% of the total biomass of brood fish stocked per pond. Water quality parameters were taken daily for dissolved oxygen and temperature; and twice weekly for pH, ammonia–N and nitrite–N. Alkalinity and hardness were recorded at the time of stocking and just before harvesting.

For the first spawning period (early), two ponds of each treatment were drained and 16 females (out of 32) were selected based on external characteristics (abdominal fullness, softness and palpability of the ovaries, redness or swollen appearance of the genitals). The second spawning period (middle) was performed using one pond of each treatment, selecting 16 females (out of 32). The last spawning period (late) selected all the remaining females in all ponds.

Hormone injections

Selected females were transferred to holding tanks (per treatment) supplied with continuous flow-through water and placed individually in soft mesh bags. Total length, body weight and girth were recorded. The LHRHa from Sigma Chemical Co (St Louis, MO, USA) was utilized. Hormone injections were administered in two doses: a priming injection of $20\,\mu\mathrm{g\,kg^{-1}}$ LHRHa, followed 12:00 hours later by a resolving dose of $100\,\mu\mathrm{g\,kg^{-1}}$.

Collection and fertilization of gametes

Twenty-four hours after the second injection, females were monitored for ovulation. Females with released eggs were removed from the holding tank and anaesthetized in $250\,\mathrm{mg\,L^{-1}}$ tricaine methane sulphonate (MS-222) (Argent Chemical Laboratories, Redmond, WA, USA). Females were then stripped and eggs were collected in aluminium containers lubricated previously with vegetable shortening. Those females that did not express eggs were returned and rechecked later. Stripping of gametes ceased when all females had been stripped or attempts to strip them had been made. Females that spawned were classified in three different ways: by the age of the broodstock females, by the period of spawning and strain either as a proportion (Table 2) or as a frequency (Table 3).

Table 2 Proportion of female channel catfish spawned (%) by age and period of spawning (early, middle and late season) after being maintained on the various dietary treatments

			Percentage spawned (%)					
Treatment	**Protein (%)**	**Feed frequency**	**Age 3**	**Age 4**	**Age 5**	**Period 1**	**Period 2**	**Period 3**
Strain 1								
1	42	6	59	33	78	79	100	39
2	42	3	35	47	72	93	67	28
3	32	6	44	29	90	79	67	54
4	32	3	41	75	100	92	67	52
Total			44	46	85	85	75	42
Strain 2								
1	42	6	44	83	73	88	50	58
2	42	3	40	50	85	93	67	65
3	32	6	36	50	63	87	71	38
4	32	3	53	100	79	100	100	54
Total			43	71	75	92	74	54

Statistical analysis

Analysis of the spawning was performed using (1) ordinary linear regression (ANOVA and ANCOVA) after the response is arcsine transformed; (2) χ^2 analysis and Fisher's exact test after the response is converted to percentages and (3) logit analysis and exact logit analysis using binary spawn–no spawn response. All the models were constructed to include the effect of treatments (protein and frequency) on spawning success for each strain and use the age of fish and period of spawning as covariates. The GLM, frequency and logistic procedures of SAS® version 9.1.3 were utilized for the analysis.

Performing ANOVA requires data transformation as percentages cannot be used directly in this type of analysis. The percentage response is arcsine transformed before proceeding with ANOVA. The statistical model for this analysis is given by

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

where μ represents the average response, α_i, $i = 1,2$, represents the effect of the ith level of protein, β_j, $j = 1,2$, represents the effect of the jth level of feeding frequency, $(\alpha\beta)_{ij}$, $i, j = 1,2$, represents the interactive effect of the ith level of protein and the jth level of feeding frequency and ε_{ijk} represents the random error associated with the arcsine-transformed response (y_{ijk}) of the kth fish that was treated with the ith level of protein and the jth level of feeding frequency.

An extension of the ANOVA approach is the ANCOVA model given by

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \eta x_{ijkl} + \varepsilon_{ijkl}$$

where γ_k, $k = 1,2,3$, is the effect of period k and η is an unknown regression coefficient to be estimated from the data. Furthermore, y_{ijkl}, x_{ijkl} and ε_{ijkl} represent the response, age and random error, respectively, of the lth fish that was treated with the ith level of protein and the jth level of feeding frequency in period k.

The χ^2 analysis tests whether the differences between the observed count and the expected count could have occurred by chance alone. Data are reported in raw frequencies (not percentages) and are assumed to be independent. Observed frequencies cannot be too small; otherwise, the assumption of normal distribution of sample frequencies will not be met. When this happens, the remedial measure involves performance of Fisher's exact test, which is a non-parametric test that does not depend on any large-sample distribution assumption. The *P*-value of Fisher's exact test is just the proportion of tables that are as extreme as or more extreme than the table given by the observed sample in all possible permutations of the data (Hollander & Wolfe 1999). In our case, we are testing for an association in a cross-classified table for each strain, where female age and period of spawning are independent variables between TC and success (or failures) in the spawning process (Table 3).

Logit analysis was performed by keeping the response in its simplest form as binary data where a response of 0 is recorded for a fish that did not spawn successfully and a response of 1 is recorded for a fish that spawned successfully. This enables one to consider the binary response corresponding to the fish in the cell (i,j,k,l) as a random variable that follows a Bernoulli's distribution with an unknown probability of success p_{ijkl}. Logit analysis proceeds by linking this probability of success to the explanatory variables through a logit link function in the spirit

Table 3 Total number of broodstock spawned on the various dietary treatments, by strain, age and period (Per.) of spawning (1, early; 2, middle; 3, late season)

			Age 3			Age 4			Age 5		
Treatment	**Protein (%)**	**Feed frequency**	**Per. 1**	**Per. 2**	**Per. 3**	**Per.1**	**Per. 2**	**Per. 3**	**Per.1**	**Per. 2**	**Per. 3**
Strain 1											
1	42	6	3 (4)	3 (3)	4 (10)	0 (1)	3 (3)	0 (5)	8 (9)	1 (1)	5 (8)
2	42	3	6 (6)	0 (1)	1 (13)	4 (5)	2 (3)	2 (9)	3 (3)	4 (5)	6 (10)
3	32	6	3 (6)	1 (2)	3 (8)	2 (2)	0 (1)	0 (4)	6 (6)	3 (3)	10 (12)
4	32	3	5 (5)	1 (2)	3 (15)	1 (2)	1 (2)	4 (4)	6 (6)	2 (2)	6 (6)
Total			17 (21)	5 (8)	11 (46)	7 (10)	6 (9)	6 (22)	23 (24)	10 (11)	27 (36)
Strain 2											
1	42	6	5 (5)	0 (2)	2 (9)	2 (2)	0	3 (4)	7 (9)	3 (4)	14 (20)
2	42	3	2 (3)	0 (1)	4 (11)	1 (1)	0	0 (1)	10 (10)	4 (5)	20 (25)
3	32	6	2 (3)	0	3 (11)	1 (1)	0	1 (3)	10 (11)	5 (7)	9 (20)
4	32	3	6 (6)	2 (2)	0 (7)	0	1 (1)	1 (1)	10 (10)	5 (5)	18 (27)
Total			15 (17)	2 (5)	9 (38)	4 (4)	1 (1)	5 (9)	37 (40)	17 (21)	61 (92)

The total number of females that were evaluated on the various dietary treatment is in parentheses.

of GLMs (McCullagh & Nelder 1999). This is given by

$$\log\left(\frac{p_{ijkl}}{1-p_{ijkl}}\right) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \eta x_{ijkl}$$

where all the model parameters are as defined above. The left-hand side of this equation is the natural logarithm of the odds of successful spawning (the ratio of the expected number of times of spawning success to the expected number of times of spawning failure), while the right-hand side of this equation is a linear function composed of treatment effects and covariates. Solving for the probability of successful spawning as a function of the treatment effects and covariates, we obtain the equivalent form given by

$$p_{ijkl} = \frac{e^{\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\gamma_k+\eta x_{ijkl}}}{1 + e^{\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\gamma_k+\eta x_{ijkl}}}$$

where e is the base of the natural log function.

Traditional inference for logit analysis uses asymptotic distribution results based on the parameters estimated using the unconditional maximum likelihood method. This method is asymptotic and hence requires a large sample size to provide a reliable inference. Our data, however, were sparse in that it had a number of cells with small observed frequency counts. This was handled by implementing the exact logistic regression procedure given by Cox (1970), Hirji, Mehta and Patel (1987) and Mehta and Patel (1995). Given n observations, the exact logistic regression proceeds by determining how likely the observed response is among the 2^n possible response permutations in a manner that mimics Fisher's exact test. Inference is based on the exact distribution of parameters of interest conditional on the remaining parameters. A direct enumeration of all the possible permutations becomes infeasible quite quickly. For instance, for our data set of 414 fish there are $2^{414} = 4.23 \times 10^{124}$ possible permutations. To make this type of exact conditional inference computationally practical, we used the multivariate shift algorithm of Hirji *et al.* (1987).

Results

ANOVA and ANCOVA

For the ANOVA model, the protein level, feed frequency and the interaction between the two were found to be non-significant (Table 4), with the exception of strain 2 (low spawning) where feed frequency was suggestive ($P = 0.0562$). The ANCOVA showed that both covariates (age of fish and period of spawning) have a highly significant effect on the proportion of females that spawned successfully for both strains. Regarding the effects of the protein level, feed frequency and their interactions in the ANCOVA model, no significant effect (Table 5) on the proportion of females that spawned successfully was found, with the exception of strain 2 (low spawning) fish, where the feed frequency was found to have a significant effect ($P = 0.0408$).

Table 4 Results of ANOVA evaluating differences in spawning success from channel catfish females, among two dietary protein levels, and two feed frequencies, for each strain

Source	d.f.	*F*	*Pr* > *F*
(a) Strain 1 (high spawning)			
Protein	1	1.64	0.2024
Feed frequency	1	0.38	0.5719
Protein × feed frequency	1	0.78	0.3796
(b) Strain 2 (low spawning)			
Protein	1	0.53	0.4820
Feed frequency	1	3.68	0.0562
Protein × feed frequency	1	0.78	0.3779

Table 5 Results of ANCOVA evaluating differences in spawning success from channel catfish females, among two dietary protein levels, and two feed frequencies, for two strains (high and low spawning) using age of fish and period of spawning as covariates

Source	d.f.	*F*	*Pr* > *F*
(a) Strain 1 (high spawning)			
Protein	1	2.14	0.1454
Feed frequency	1	0.02	0.8880
Protein × feed frequency	1	1.59	0.2094
Age of fish	1	28.01	<0.0001
Period of spawning	2	17.61	<0.0001
(b) Strain 2 (low spawning)			
Protein	1	1.18	0.2786
Feed frequency	1	4.23	0.0408
Protein × feed frequency	1	0.88	0.3499
Age of fish	1	23.11	<0.0001
Period of spawning	2	17.62	<0.0001

Pearson's χ^2 and Fisher's exact test

We were unable to use the Pearson χ^2 analysis due to low-frequency counts in the cells (five or less). Fisher's exact test was used as a remedial measure. Analysis revealed that the TC had a significant effect ($P = 0.0041$) on spawning success only for strain 1, 4-year-old fish that were spawned in period 3 (late) (Table 6).

Logit analysis and exact logit analysis

Table 6 Results of Fisher's exact test evaluating differences in spawning success from channel catfish females, among two dietary protein levels, and two feed frequencies, for two strains (high and low spawning) using age of fish and period of spawning as variables

	Strain 1		Strain 2	
	Table probability	*Pr* ≤ *P*	Table probability	*Pr* ≤ *P*
Age 3				
Period 1	0.0134	0.0809	0.0662	0.1103
Period 2	0.0714	0.3571	0.1000	0.2000
Period 3	0.0052	0.2415	0.0120	0.3876
Age 4				
Period 1	0.0833	0.5000	N/A	N/A
Period 2	0.0714	0.4643	N/A	N/A
Period 3	0.0005	0.0041	0.0952	0.5714
Age 5				
Period 1	0.3750	1.0000	0.0401	0.1624
Period 2	0.4545	1.0000	0.0702	0.7953
Period 3	0.0082	0.2516	0.0005	0.1050

N/A, not applicable.

Once again, the low-frequency counts make the performance of the traditional logistic regression analysis unsatisfactory. Thus, we performed the exact logistic analysis and the results obtained with this analysis are presented in Table 7. For strain 1, we found no significant effect of protein level, feed frequency or their interaction on the probability of successful spawning. For strain 2, only feed frequency had a significant effect ($P = 0.0427$) on the probability of successful spawning. Considering the covariates, the age of the fish had a highly significant effect ($P < 0.0001$ in both strains), period 1 (early) had a significant effect ($P = 0.0022$ for strain 1 and $P < 0.0001$ for strain 2), and period 2 had no significant effect ($P = 0.9533$ for strain 1 and $P = 0.7776$ for strain 2) on the probability of successful spawning.

For strain 2, the odds of successful spawning of fish that are fed three times per week were estimated to be 1.9 times that of fish fed six times per week. Furthermore, for strain 1, the odds of successful spawning of female 5-year-old fish were estimated to be 8.4 times higher than female 3-year-old fish and 2.9 times higher than female 4-year-old fish. For strain 2, the odds of successful spawning of female 5-year-old fish were estimated to be 4.9 times higher than female 3-year-old fish and 2.2 times higher than female 4-year-old fish. Finally, the estimated odds of successful spawning during period 1 (early) were 8.6 and 16.3 times higher than the odds in period 3 (late) for strains 1 and 2 respectively, while no significant difference in the probabilities of spawning was detected between period 2 (middle) and period 3 (late).

Discussion

Logit analysis was developed as an alternative approach to ANOVA and ANCOVA. Conceptually, this

Table 7 Results of logistic regression analysis evaluating differences in spawning success from channel catfish females, among two dietary protein levels, and two feed frequencies, for two strains (high and low spawning) using age of fish and period of spawning as covariates

Variable*	d.f.	Parameter estimate	*Pr* > chi-square	Odds ratio
(a) Logistic regression analysis – strain 1				
Protein	1	0.2595	0.1215	
Feed frequency	1	0.0399	0.8619	
Protein × feed frequency	1	0.2679	0.1184	
Age	1	1.0268	<0.0001	2.8
Period 1	1	1.0449	0.0022	8.6
Period 2	1	0.1936	0.9533	
(b) Logistic regression analysis – strain 2				
Protein	1	− 0.1751	0.2707	
Feed frequency	1	0.3150	0.0427	1.9
Protein × feed frequency	1	0.1213	0.4312	
Age	1	0.7988	<0.0001	2.2
Period 1	1	1.3963	<0.0001	16.3
Period 2	1	− 0.2734	0.7776	

*Period 1 and period 2 correspond to the logistic regression coefficients where the levels of period are coded as period 1 (1, 0)–period 2 (0, 1)–period 3 (− 1, 1).

analysis makes more sense and has better statistical properties. It provides an optimal method for regression analysis of dichotomous-dependent variables (Allison 1999). In addition to clear-cut violations of assumptions, ANOVA and ANCOVA use an arcsine transformation that makes the interpretation of regression coefficients difficult as they are not estimated on the original scale (Byers *et al.* 2003). Logit analysis has previously been effectively utilized in fisheries to analyse fish maturation among other things. These studies used asymptotic distributional results that do not hold in our case due to small cell frequencies. Exact logit analysis was used to provide correct inference as recommended by Stokes *et al.* (2000).

In our exact logit analysis, the effect of protein on spawning success, high protein levels (42%) versus low protein levels (32%), was not statistically significant for both strains (Table 7). Similar results were found for rainbow trout, *Salmo gairdneri* (Richardson), where fish fed at 27%, 37%, 47% and 56% protein levels did not show a significantly different effect on the success of spawning (Roley 1983). Feed frequency had a significant effect on spawning success only in strain 2 (low spawning), where fish fed every other day (three times per week) had 1.9 times more odds of spawning than those fed daily (six times per week). Catfish farmers seem to be aware of this situation according to Steeby and Wagner (2005), who reported that about 23.5% of broodstock breeding operations feed their animals daily during spring/early summer, while 68.8% feed every other day or every third day (35.7% every other day and 33.1% every third day). Hence, the results of the exact logit analysis match the expected biological response and would thus be considered a more appropriate test.

Age of fish is often considered to be a determinant factor in spawning performance. Santiago (1979) reported a very low spawning success in channel catfish 3-year-old females (12.7%), and our study shows that using 5-year-old female fish rather than younger females is likely to produce a significantly higher rate of successful spawning. A similar conclusion was reached by Steeby and Wagner (2005) who wrote 'considering the time and capital investment represented by broodfish and the fact that good egg output is expected from fish three years and older – producers should use brooders until the fish are at least five years old'. Although older fish performed better, producers must carefully consider other factors, such as economic and genetic, along with the results presented in this paper in making their decision.

Results of this case study demonstrate that appropriate statistical models provide better insight into the cause–effect relationships that exist between control variables and the dependent variable (likelihood of spawning in this case). For spawning success or even other dichotomous variables (such as survival–mortality and infected–not infected), logistic and other binary regression models are generally considered to be better statistical tools as they offer detailed descriptions of the relationships among variables in addition to their theoretical optimality. Using exact logit analysis, channel catfish broodstock females from Auburn University, classified as strain 1 (high spawning) and strain 2 (low spawning), were found to exhibit significantly different spawning success probabilities for different levels of feed frequency but not for different levels of protein. Strain 2 fish fed six times per week had a significantly lower success in spawning as compared with strain 2 fish fed three times per week. Whereas feeding frequency did not have a significant effect on the spawning success of strain 1 fish, age of fish and period of spawning also had a significant effect on the probability of spawning success.

In summary, as would be expected, each strain of fish responded somewhat differently to the test variables. Changing the protein level of the diet from 32% to 42% or increasing the feeding frequency from three to six times per week either did not influence spawning or negatively affected spawning respectively. Additionally, older fish performed better than younger fish and the early spawning period was better than the later spawning period, regardless of strain. These responses were only detected using logit analysis, which was a more sensitive test and would thus be recommended for this type of data.

Acknowledgments

The authors would like to thank those who have taken the time to critically review this manuscript as well as those who helped in supporting this research. The Southern Regional Aquaculture Center SRAC, Grants No. 2001-38500-10307 and 2002-38500-11805 supported this research. Mention of a trademark or proprietary product does not constitute an endorsement of the product by Auburn University and does not imply its approval to the exclusion of other products that may also be suitable.

References

Allison P.D. (1999) *Logistic Regression Using SAS® System: Theory and Application*. SAS Institute, Cary, NC, USA, 288pp.

Bernardo J. (1993) Determinants of maturation in animals. *Trends in Ecology and Evolution* **8**, 166–173.

Bondari K., Ware G.O., Mullinix B.G. Jr & Joyce J.A. (1985) Influence of brood fish size on thee breeding performance of channel catfish. *The Progressive Fish-Culturist* **47**, 21–26.

Bromage N. (1995) Broodstock management and seed quality – general considerations. In: *Broodstock Management and Egg and Larval Quality* (ed. by N.R. Bromage & R.J. Roberts), pp. 1–24. Blackwell Science, Cambridge, UK.

Byers A.L., Allore H., Gill T.M. & Peduzzi P.N. (2003) Application of negative binomial modeling for discrete outcomes: a case study in aging research. *Journal of Clinical Epidemiology* **56**, 559–564.

Celada J.D., Antolin J.I., Carral J.M., Pérez J.R. & Sáez-Royuela M. (2006) Reproductive efficiency of the signal crayfish (*Pacifastacus leniusculus* Dana. Ascacidae) at different densities under both culture and laboratory conditions. *Aquaculture* **252**, 298–304.

Cox D.R. (1970) *Analysis of Binary Data*. Methuen, London, UK, 142pp.

DeMartini E.E. & Lau B.B. (1999) Morphometric criteria for estimating sexual maturity in two snappers, *Etelis carbunculus* and *Pristipomoides sieboldii*. *Fishery Bulletin* **97**, 449–458.

Donaldson E.M. (1996) Manipulation of reproduction in farmed fish. *Animal Reproduction Science* **42**, 381–392.

Friendly M. (2000) *Visualizing Categorical Data*. SAS Institute, Cary, NC, USA, 456pp.

Heino M., Dieckmann U. & Godø O.R. (2002) Measuring probabilistic reaction norms for age and size at maturation. *Evolution* **56**, 669–678.

Hirji K.F., Mehta C.R. & Patel N.R. (1987) Computing distributions for exact logistic regression. *Journal of the American Statistical Association* **82**, 1110–1117.

Hollander M. & Wolfe D.A. (1999) *Nonparametric Statistical Methods*, 2nd edn. John Wiley & Sons, New York, USA, 787pp.

Keating K.A. & Cherry S. (2004) Use and interpretation of logistic regression in habitat-selection studies. *Journal of Wildlife Management* **68**, 774–789.

Kellogg R.L., Ligotino R.J. & Jinks S.M. (1984) Thermal mortality prediction equations for entrainable striped bass. *Transactions of the American Fisheries Society* **113**, 794–802.

Lambert D.M. (1998) *Comparison of techniques for artificial production of channel catfish (Ictalurus punctatus females) × blue catfish (I. furcatus males) hybrid embryos*. MS thesis. Department of Fisheries and Allied Aquacultures, Auburn University, Auburn, AL, USA. 81pp.

Mehta C.R. & Patel N.R. (1995) Exact logistic regression: theory and examples. *Statistics in Medicine* **14**, 2143–2160.

McCullagh P. & Nelder J.A. (1999) *Generalized Linear Models*. Chapman & Hall/CRC, London, UK/Boca Raton, FL, USA, 511pp.

McHugh P. & Budy P. (2004) Patterns of spawning habitat selection and suitability for two populations of spring chinook salmon, with an evaluation of generic versus site-specific suitability criteria. *Transactions of the American Fisheries Society* **133**, 89–97.

Munger C.R., Wilde G.R. & Follis B.J. (1994) Flathead catfish age and size at maturation in Texas. *North American Journal of Fisheries Management* **14**, 403–408.

Peck M.A. & Holste L (2006) Effects of salinity, photoperiod and adult stocking density on egg production and egg hatching succes in *Acartia tonsa* (Calanoida: Copepoda): optimizing intensive cultures. *Aquaculture* **255**, 341–350.

Rodríguez-González H., García-Ulloa M., Hernández-Llamas A. & Villareal H. (2006) Effect of dietary protein level on spawning and egg quality of redclaw crayfish *Cherax quadricarinatus*. *Aquaculture* **257**, 412–419.

Roley D.D. (1983) *The effect of diet protein level, feeding level and rearing water temperature on the growth and reproductive performance of rainbow trout broodstock*. PhD dissertation. University of Washington, Seattle, WA, USA, 271pp.

Santiago A.C. (1979) *Effects of feeding regime on reproductive performance of female channel catfish in ponds*. PhD dissertation. Department of Fisheries and Allied Aquacultures, Auburn University, Auburn, AL, USA, 74pp.

Steeby J.A. & Wagner B.A. (2005) Channel catfish hatchery and fry production practices in the U.S. catfish industry. *World Aquaculture* **36**, 14–17.

Stokes M.E., Davis C.S. & Koch G.G. (2000) *Categorical Data Analysis Using the SAS System*. SAS Institute, Cary, NC, USA, 626pp.

Zelterman D. (1999) *Models for Discrete Data*. Oxford University Press, New York, USA, 233pp.

Zelterman D. (2002) *Advanced Log-linear models using SAS®*. SAS Institute, Cary, NC, USA, 189pp.