Analysis of recreational fish catches – dealing with highly skewed distributions with many zeros

Michael F. O'Neill[1] and Malcolm J. Faddy[2]

[1] Southern Fisheries Centre, Department of Primary Industries, Deception Bay, Queensland 4508, Australia
[2] School of Mathematics and Statistics, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Abstract

Data from surveys of recreational anglers fishing on three estuaries in eastern Australia reveal highly skewed distributions of catches with many zeros. Such data may be analysed using a two component approach involving a binary (zero/non-zero catch) response and the non-zero catches. A truncated regression model was effective in analysing the non-zero catches. Covariates were incorporated in the modelling, and their critical assessment has led to improved measures of fishing effort for this recreational fishery.

Introduction

Analysis of recreational catch data can present problems because many fishing trips fail to catch any fish (e.g. Figure 1). The resulting data are generally highly skewed and even after various transformations may not meet the assumptions required for many standard statistical techniques. For example, the use of a log transformation and normal residual distribution (Robins et al., 1998) fails to take account of the discrete nature of catch data and this becomes more problematic when the catches are small which is fairly typical in recreational fishing surveys (e.g. Figure 1). Models based on discrete distributions such as the Poisson and negative binomial fit into the generalised linear modelling framework (McCullaugh and Nelder, 1989) and use appropriate residual distributions. However, the factors influencing zero catches may well be different from those that influence non-zero catches; for example, recreational fishers may be less enthusiastic while they are not catching fish and if no fish are caught in a relatively short period of time they may go elsewhere.

Welsh et al. (1996) presented models for dealing with discrete species abundance data that contain many zeros. The models represent extensions of the Poisson and negative binomial distributions to allow for extra zeros. Such models are essential for hypothesis testing given the properties of recreational survey data with many zero values: the importance of factors affecting the recreational catch may be under- or overstated if models of this type are not used, leading to unreliable inferences. The model used here has a separate component for the zeros; this allows possibly different factors to influence this component of the model, and also allows the zero and non-zero catch data to be analysed separately.

Data on recreational catches of yellowfin bream (*Acanthopagrus australis*) from three estuaries in south-east Queensland, Australia, are used in this paper. A truncated regression model that allows for extra zeros is presented along with an account of the results.

Modelling

There are two components to the modelling and data analysis. The first component refers to the binary response of zero or non-zero catch, with the capture of fish of a species by a fishing group occurring according to the probabilities

$P_{(\text{non-zero catch})} = p$ and $P_{(\text{zero catch})} = 1 - p$

The logit of the probability p was modelled as a linear function of the covariates: estuary, season, day type, fishing platform, number of anglers in the group, fishing time and number of fishing lines in the group (the first four of these being factors).

The second component was for only those catches where a non-zero number of fish was caught. Truncating discrete distributions by conditioning on the catch being greater than zero will provide appropriate distributions to analyse these data. The Poisson and negative binomial distributions are special cases from general discrete distribution modelling described in Faddy (1997) and referred to as extended Poisson process modelling (EPPM) with transition rates:

$$\lambda_n = \log(1+\delta\mu)\left(\frac{1}{\delta} + n + \frac{c}{d+n}\right)$$

for $n = 0, 1, 2, \ldots$ (1)

where the parameter μ can be modelled as a linear function of any covariates using a log link. Here the limit as $\delta \to 0$ results in the Poisson distribution with mean μ and $c = 0$ gives negative binomial distributions with mean μ, with other values of c giving more general dispersion properties. Faddy (1997) and O'Neill (2002) describe how the transition rates (1) are used to calculate the corresponding discrete probability distribution $p_0, p_1, p_2, \ldots$. These probabilities can then be truncated to give an appropriate distribution on 1, 2, 3, … :

$$\frac{p_i}{1-p_0}, \text{ for } i = 1, 2, 3, \ldots \quad (2).$$

Data

Recreational fish catch data from roving creel surveys were collected between June 1997 and August 1998 from the Burnett River, Maroochy River and Pumicestone Passage in south-east Queensland, Australia. Overall, a minimum of five weekdays and five weekend days or public holidays, selected at random, were surveyed each month in each estuary. These days were surveyed either in a morning shift (6 am to 12 noon) or afternoon shift (12 noon to 6 pm).

The three estuaries were stratified into smaller areas to enable angler numbers to be counted. Counts were recorded on each survey shift in each area at random times. During a shift, staff would drive their boat to an area and count the number of boats and people actively fishing (with a line in the water). Once the count was complete, boat/shore fishing groups were randomly interviewed for a one-hour period. The number of persons fishing, actual fishing time (hours), number of fishing lines used, number and species of fish released, and number and size (total length in centimetres) of each fish retained were recorded for each fishing group.

Anglers from five to eight randomly selected areas were interviewed in each shift. If no anglers were present in a scheduled interview area, a zero count was recorded and another nearby area was surveyed.

Results

Data

Shown in Figure 1 is the histogram of the observed yellowfin bream catches per fishing group. High frequencies of zero catches are apparent along with considerable skewness in the upper tail of the non-zero catches.

Logistic regression of binary (zero/non - zero catch) response

Catches of yellowfin bream showed a low proportion of fishing groups actually catching fish. These proportions changed significantly with the estuary fished, the time of year and fishing platform. Also, they were dependent on the number of people in the fishing group, the time spent fishing and the number of fishing lines used. The probability of a fishing group catching yellowfin bream increased the longer they fished and the more fishing lines used. However, the probability of boat and shore groups catching yellowfin bream decreased with the number of anglers in the group, although this effect for shore groups was the least significant (p-value ≈ 0.03).

EPPM regression of non-zero catches

Both the Poisson and negative binomial based models tended to underestimate the residual variation, with the EPPM (1) and (2) doing better with an estimate of the parameter c of 0.12. In each estuary the average catch of yellowfin bream was generally less than one fish per group hour. There were significant differences in average yellowfin bream catches due to some of the variables. Catches of yellowfin bream in the Maroochy River and Pumicestone Passage were significantly higher than in the Burnett River. Average catches of yellowfin bream were highest during the winter

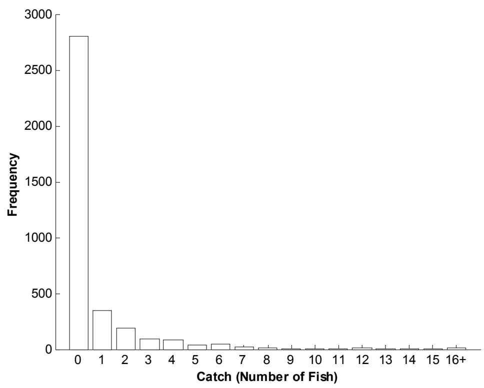

Figure 1. Observed distribution of yellowfin bream catches.

months. There were no significant differences in average yellowfin bream catches between boat and shore fishing groups, and between weekend and weekdays. There was a significant positive effect of the time spent fishing on the average catch. For shore based fishing, there was a negative relationship between average catch and the number of anglers per group, but this was the least significant effect (p-value ≈ 0.03). There was also no significant effect of the number of fishing lines in the group on average catches.

Discussion

In this paper models generalising those described in Welsh et al. (1996) have been used to analyse data on recreational fish catches. The methodology was particularly applicable to these data which exhibited many zero values and low non-zero catch sizes, as the models used more accurately reflected these properties of the data than more standard modelling options available in most statistical packages. The truncated EPPM component has adequately allowed for the considerable dispersion shown in the data (Figure 1), and the overall analysis facilitated critical assessment of important effects on recreational catches, thereby making more effective use of the survey data.

The analyses identified important factors affecting the recreational catch. Total catch for yellowfin bream should be estimated separately in each estuary, season and fishing platform. However, the catch data could be grouped across weekend and weekdays to estimate total catch. The models also indicated some interesting relationships between catch and fishing effort. As expected, for both boat and shore fishing groups, the average catch increased as the time fished increased. However, larger boat fishing groups were less likely to catch yellowfin bream than similar sized shore groups. This negative relationship probably indicated that the more serious and experienced boat anglers tended to fish by themselves or in small groups. Larger sized groups fishing from a boat may have fished more as a social activity and were therefore less likely to catch fish. Also, more fishing lines used by a given number of anglers tended to increase the likelihood of catching fish. This latter positive influence had a counteracting effect on the negative influence of larger numbers of boat anglers reducing the chances of catching fish. Overall, the results indicate that number of hours fished per group is a fair representation of boat-fishing effort, while the number of hours fished per line or angler (since these will be correlated) represents shore-fishing effort. With these measures of fishing effort and the above stratification by estuary, season and fishing platform, more reliable estimates of recreational catch rates and hence total catch can be made (O'Neill, 2000), thus providing better information for management of the fishery.

References

Faddy MJ, 1997. On extending the negative binomial distribution, and the number of weekly winners of the UK national lottery. *Mathematical Scientist* **22**:77–82.

McCullagh P, Nelder JA, 1989. *Generalized Linear Models, Second Edition.* Chapman and Hall, London.

O'Neill MF, 2000. *Fishery Assessment of the Burnett River, Maroochy River and Pumicestone Passage.* Project report QO99012. Queensland Department of Primary Industries, Australia.

—— 2002. Use of Binary and Truncated Regression Models in the Analysis of Recreational Fish Catches. Masters Thesis, The University Of Queensland, Australia.

Robins CM, Wang Y-G and Die DJ, 1998. The impact of global positioning systems and plotters on fishing power in the northern prawn fishery, Australia. *Canadian Journal of Fisheries and Aquatic Sciences* **55**:1645–1651.

Welsh AH, Cunningham RB, Donnelly CF and Linden-mayer DB, 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling* **88**:297–308.