DESIGN AND ANALYSIS OF A SOCIOECONOMIC SURVEY OF MARINE RECREATIONAL FISHING

Richard Bolstein, George Mason University
Robert L. Hiett, Kathryn A. Chandler and Audrey K. Reniere, KCA Research Inc.

1. INTRODUCTION

Since 1978 the National Marine Fisheries Service (NMFS) has been pursuing the mandate of the Fisheries Conservation and Management Act (FCMA) of 1976 through a data collection effort known as the National Statistical Survey of Marine Recreational Fishermen (NSS) [2,3]. The objective of the NSS is to obtain catch and effort data for recreational saltwater fishermen on a species basis. It attempts to estimate such quantities as total catch, weight of catch, total number of participants, and total number of fishing trips over a one-year period. While this information is necessary, it does not fully meet the provisions of the FCMA which require fisheries managers to take into account social, economic and ecological factors in developing Optimum Sustained Yield estimates and in allocating stocks among competing user groups.

During 1981 a major effort took place to obtain social and economic information from marine recreational fishermen along the Atlantic, Gulf, and Pacific coasts. Known as the Socioeconomic Survey (SES), this effort included an on-site survey with interviews from more than 7,000 fishermen. The objective of the survey was to collect information of three general types:

(a) Information on anglers (fishermen) as individuals: what type of people, from a socioeconomic standpoint, go fishing, and with what intensity?
(b) Information on the fishing experience: why and how do people fish?
(c) Trip-specific information, such as expenditures, disposition of catch by species, levels of satisfaction and factors affecting such.

The purpose of this paper is to present the main aspects of the sample design, data collection methods, and analysis techniques used in this survey as performed by KCA Research, Inc., under a grant from the National Marine Fisheries Service.

2. SAMPLE DESIGN

It was apparent at the beginning of this study that some of this information could not be obtained during an on-site interview. In particular, actual disposition of fish and total expenditures per trip would not be known until after the fishing trip was completed. Therefore, the on-site survey was complemented by a followup telephone survey to the fishermen to collect the full set of data needed for the study. A separate survey was undertaken for each coast. Results were reported separately for each coast but not overall since there was no basis for comparison among coasts (species are different, as well as climate and geography). We refer only to the Atlantic Coast in this paper, but the methodology is similar for the others.

Three principal variables that affect the fishing experience are:

(1) Time of year
(2) Geographical location
(3) Mode of fishing

Mode was classified into four types: fishing from man-made structures (piers, docks, jetties, or bridges), natural shore areas, party or charter boats, and private or rental boats. The Atlantic Coast was divided into three subregions corresponding to different fishing characteristics (such as different species habitats). Time of year was divided into six two month periods called waves.

The three variables mode, wave, and subregion were used as stratifying variables. Thus, the universe for the Atlantic Coast was decomposed into 72 (4 x 6 x 3) strata. The unit of analysis was the fishing trip (not fisherman). Sample size for the Atlantic Coast was set at 3600 by NMFS. The sample was allocated among the 72 strata based on estimates of the number of trips in each strata computed from the 1979 NSS Survey ([3]). The allocation was roughly proportional, except for the party/charter boat mode which was increased by mandate from NMFS. For the data analysis however, estimates of the number of trips in each strata were taken from the 1981 NSS Survey, which was being pursued independently and simultaneously with the SES Survey. Thus, in the final analysis, the stratified sample was not proportional.

To select the sampling units, fishing sites were selected by simple random sampling (without replacement within wave) from a site list by state and mode. A random date within wave was selected for each chosen site, except that 90% of the dates were allocated to weekends. An interviewer was given a quota of interviews to obtain on that date for that mode (in some cases, two modes) using the selected site as the starting site. If the quota could not be met, the interviewer selected an alternative site within the same county where he (she) anticipated additional interviews could be obtained. A third site was chosen if necessary. Over-weighting weekends helped to reduce the number of alternative sites needed.

Thus, the survey design was a stratified sample with fishing sites serving as sampling units and fishing trips serving as units of analysis. Rather than regard the selection within strata as a three stage sample (sites, dates, trips) it is more convenient to treat this as a two-stage sample with site-days serving as the primary sampling units. Note that these PSU's were not selected by simple random sampling since no site could be chosen on two different dates within the wave, and, of course, since weekends were overweighted. We will argue later, however, that at least for many variables of importance in the study srs can be assumed for the PSU's within strata.

3. DATA COLLECTION

Interviewers trained in fish identification skills (primarily recruited from coastal universities) performed the intercept part of the

survey. They were instructed to conduct an interview after a fisherman had completed fishing for the day, i.e. completed his/her fishing trip. This avoided the bias of oversampling 'longer trips' which would have been the case had fishermen been interviewed during rather than after their trips (see the discussion in [2]). Interviewers used their own judgment as to what time to arrive at the starting site to obtain the maximum number of interviews, usually in early or mid-morning.

Information obtained during the on-site interview included the number of fish caught by species, the number of these which were thrown back or otherwise not kept and the reasons why, fishing technique used, and the name of the target species the fisherman had for the trip, if any. Questions relating to the social aspects of fishing were also asked, such as reasons for fishing, whether fishing with friends, family, business associates or other types of groups, resident/visitor status, and distance traveled to site. Each fisherman was also asked how many trips he had taken in the past 12 months from the same site in the same mode. Finally, each angler was asked if he/she would participate in a telephone follow-up interview. Approximately 2800 of the 3600 intercepted on the Atlantic Coast agreed to a follow-up.

The telephone follow-ups were conducted within a mean time of 18.8 days after the original intercept. About 2200 were able to be contacted. The follow-up telephone survey employed a technique known as Computer Assisted Telephone Interviewing. As each questionnaire came in from the field, it was edited and keypunched. A computer file was generated and telephone interviewers were then able to access each respondent's answers given during the on-site interview to obtain the follow-up data. Thus it was possible to remind each respondent of the number of each species of fish he had caught while questions were asked about disposition of fish, expenditures, and level of satisfaction with the trip. Data was then entered directly into computer terminals with edit checking routines included.

Fisherman-specific information was also obtained on the telephone follow-up, such as approximate number of days fished during the past 12 months, which we call avidity, number of years fished, occupational status, and household income.

4. ESTIMATION WITHIN STRATA

Once a site-day was selected within a mode, wave, and subregion, an interviewer attempted to obtain the quota of interviews assigned to the site. Although the interviewer selected anglers to question, every attempt was made to minimize bias. Interviewers were instructed to approach 'unfriendly' faces, as well as those eager to talk. They were trained to take systematic samples in cases where the number of anglers exceeded the quota. For example, an interviewer might board a charter boat and systematically select anglers at their stations as early as possible during the trip.

If the quotas were always met at one site, we could regard each site as a cluster of trips from which a subsample was taken. It would be reasonable to treat this subsample as simple random sampling in view of the remarks just made. However, the quotas were often not met at the intitial site, in which case the interviewer proceeded to a nearby site as described in Section 2. Since the fishing experience should be very similar at nearby sites on the same day and in the same mode, it is reasonable to associate trips with the initial site even if they came from nearby sites on the same day. Hence, we associate a cluster of trips with each initial site-day i, from which we draw a simple random sample of size t_i. Let T_i be the actual number of trips in the population associated with site-day i (T_i includes the total number of trips taken at site i and all nearby sites visited on that day). The ratio-to-size estimate of the mean $\bar{Y}_h$ of a variate y in stratum h (see [1, section 11.8]) is

$$(1) \qquad \hat{\bar{Y}}_h = (\sum T_i \bar{y}_i)/(\sum T_i)$$

where $\bar{y}_i$ is the sample mean at site-day i and the sums are taken over all sampled site-days i in stratum h. This is a biased estimator with mean square error estimated by

$$(2) \qquad v[\hat{\bar{Y}}_h] \doteq [\sum T_i^2 (\bar{y}_i - \hat{\bar{Y}}_h)^2]/[\bar{T}_h^2 u_h (u_h - 1)] + [\sum T_i^2 (1-f_i) s_i^2 / t_i]/[u_h U_h \bar{T}_h^2]$$

where the sums are taken over all sampled i in stratum h,

U_h=total number of clusters in stratum h,

u_h=total number of clusters sampled in stratum h,

$\bar{T}_h$ =mean size of the clusters in stratum h (which is estimated by the mean size of the sampled clusters),

s_i^2 = sample variance of y at site-day i, and

$f_i = t_i/T_i$

The trouble with this estimator is that the number of trips M_i associated with site-day i is not usually known. In many cases $t_i = T_i$, especially in the off-season waves. However, in the prime fishing season T_i often exceeded the quota. Usually, interviewers also exceeded the quota in these times to make up for unfulfilled ones in the off-season, bringing t_i a little closer to T_i. However, it was not possible for them to determine T_i in general. One way out is to disregard unsampled trips at the site, that is, assume $t_i = T_i$ for all i. This amounts to a single-stage cluster sample within each strata. In this case, (1) reduces to the ordinary sample mean of stratum h, and the second term of (2), the within-site variance, vanishes. (Ghosh claims in [2], although without proof, that the resulting variance can be approximated by srs, that is, the clustering effect is neglibible if the size and mean of the clusters are independent. We discuss this in another paper.) An alternative method would be to provide a good estimate of T_i. This was not done due to lack of information. This is an area for further study if the survey is to be performed again.

5. ESTIMATION OF STRATA SIZES.

Estimates of the number of fishing trips in each strata were obtained from the 1981 NSS Survey which was conducted independently by Market Facts, Inc. with KCA Research, Inc. as subcontractor. The survey was a complemented design of a household telephone survey coupled with an independent on-site intercept survey (see [2] for a discussion of the 1979 Survey). In the telephone survey, a simple random sample of households located in the 'coastal zone' of each state was taken. The 'coastal zone' was comprised of counties, or portions thereof, within a certain distance (about 25 miles) of the coast. The number of fishermen in a sample household and the number of fishing trips they took in the past two months in their state was obtained by mode (along with other information of no immediate concern to us here). From this information, the mean number of in-state fishing trips per coastal household in the sample was computed for each state by mode and wave. Multiplying by the total number of households in the coastal zone yielded estimates of the total number of fishing trips taken by coastal zone residents in their state by mode and wave.

This estimate was expanded to yield estimates of the total numbers of fishing trips in each state using results of the intercept survey. In the latter survey, fishing trips were selected at random in each state by the same procedure as in the SES Survey. From the addresses of intercepted fishermen, the number n of sampled trips in each state made by other than coastal zone residents of the same state as well as the number c of trips made by coastal zone residents of that state were obtained by mode and wave. The estimates of trips taken by coastal zone residents was expanded by the factor n/c to provide estimates of the total number of trips by mode and wave in each coastal state. Some difficulties arose with this procedure since in some wave-mode categories the number of intercepts was too small to make the ratio n/c reliable, or even worse, either n or c was zero. In such cases, states were combined into subregions and waves were combined to obtain more reliable expansion factors.

These state estimates were combined to estimate the number of fishing trips in each strata of the SES survey.

6. OVERALL ESTIMATES

Although weekends were overweighted when assigning dates to sites (Section 2), the intraday variation should be negligible for trip-specific variables relating to catch and expenditures. In these cases we can regard the PSU's as selected by srs and use

$$(3) \qquad \bar{y}_{st} = \sum (N_h/N)\bar{y}_h \quad , \text{ and}$$

$$(4) \qquad v[\bar{y}_{st}] = \sum_h (N_h/N)^2 v[\bar{y}_h]$$

as estimates of the population mean and variance of $\bar{y}_{st}$ respectively, where

N_h = number of trips in stratum h (as estimated in Section 5)

$\bar{y}_h$ = ordinary sample mean in stratum h (from (1) with $t_i = T_i$)

For variables relating to anglers themselves such as 'occupational status', and to the fishing experience such as 'type of fishing party' these formulas may produce biased results. For example, a greater percentage of retired persons fishing alone would be intercepted on weekdays than on weekends. Formulas (3) and (4) would be adjusted by appropriate weight factors in these cases.

7. COMPUTATIONAL METHODS AND SPSS

SPSS was used for all computations. It does not have a procedure to simply compute stratified means and variances. Separate runs were required to produce within-strata means and variances and to combine them into overall estimates. Moreover, most variables had to be treated separately since the number of respondents varied, sometimes considerably, from question to question (primarily due to 'skip' patterns in the survey instrument) and would necessitate a large raw output if variables were lumped. The programming became even more complex in the case of ratio estimates, such as the percent of landed fish that were kept. Since the study involved several hundred variables, it was too costly to produce standard errors except for a few variables. However, standard errors are not usually of interest in a study of this kind except for the key variables (along with an upper bound for the others), although means, etc. are required for all items.

Fortunately, the case weighting procedure available in SPSS ([5, Chapter 9]) enabled us to compute stratified means and investigate relations between many variables efficientl using standard procedures (such as BREAKDOWN and CROSSTABS). We now describe the weighting procedure and how it was used in the study.

Consider the case where it is desired to estimate the mean of a continuous response variable, such as mean distance traveled to site. Let n_h denote the number of observations in stratum h producing a valid value of the variable. The correct estimate (3) would not be produced by SPSS CONDESCRIPTIVE since SPSS treats all samples as simple random samples. However, if the cases are weighted by $w_h = KN_h/n_h$, where K is an arbitrary constant, then the weighted sample has KN_h cases in stratum h, so the CONDESCRIPTIVE Procedure will compute the mean

$$(5) \qquad \hat{\bar{y}} = [\sum_h \sum_{i=1}^{n_h} (K\, N_h/n_h)\, y_{hi}]/[\sum_h K\, N_h]$$

$$= [\sum_h N_h \bar{y}_h]/N = \bar{y}_{st} \ ,$$

where y_{hi}, i = 1, 2, ..., n_h are the values of y for the sample units in stratum h.

Thus, SPSS Procedure CONDESCRIPTIVE produces the correct mean when acting on the weighted sample. This assumes that the number n_h of valid responses for a particular variable is known so the weight factors w_h can be computed. One may compute n_h in a separate preliminary run for the variable of interest, but this is laborious when there are a large number of variables. If instead we weight by

$$w_h = K\, N_h/m_h$$

where m_h is the sample size of strata h, then (5) becomes

$$\hat{\bar{y}} = [\sum_h \sum_{i=1}^{n_h} (K N_h/m_h) y_{hi}]/[\sum_h (K N_h/m_h) n_h] \text{ , or}$$

$$(6) \qquad \hat{\bar{y}} = [\sum_h N_h (n_h/m_h) \bar{y}_h]/[\sum_h N_h (n_h/m_h)].$$

Observe from (6) that $\hat{\bar{y}} = \bar{y}_{st}$ if n_h/m_h is a constant independent of h. Thus, the correct estimate of the mean is produced if there are either no nonrespondents ($n_h = m_h$) to the particular question at hand, or if the number of nonrespondents in each stratum is proportional to stratum size. The choice of sample size m_h to use for weighting can be varied for different groups of variables. For example, the telephone followup sample was smaller than the intercept sample, so the weighting factors would be different when estimating means of variables arising in the two different parts of the survey. Thus, several different weighted samples must be used to make the estimates.

The same logic applies when comparing means of a continuous variable across values of a categorical control variable (say, distance traveled to fishing site by target species). In this case the SPSS BREAKDOWN procedure can be applied to the appropriate weighted sample.

Finally, we note that the correct standard errors cannot be produced by the weighting method.

8. MAIN RESULTS

A great deal can be said about the recreational fishing experience from this study. Whereas results pertaining to fishing trips are on solid ground, results about fishermen are biased towards those who fish more frequently. This follows since fishing trips, and not fishermen, were sampled. Thus, a highly avid fisherman had a higher probability of being sampled than an occasional fisherman. To illustrate this point, the mean number of trips taken in the past year by fishermen in the sample was 23.5 days for the Atlantic Coast, whereas a random sample of fishing households in the coastal zone taken from the NSS survey produced a mean of only 17.2 days. It is not possible to correct for this bias by weighting since information was not collected on how many trips a fisherman took in each strata. This is but one indication of how the survey can be improved in the future. With this caveat we mention some of the main conclusions of the study for the Atlantic Coast.

Fishing is a male-oriented activity in which persons of all racial and ethnic groups, age groups, and income groups participate. It is an unorganized group activity involving family and friends. Retired people who fish do so more frequently than the general fisherman. Fishing is a relatively expensive activity with expenditures per trip averaging $59 in the party/charter boat mode, $38 for private/rental boat, and about $25 in the non-boat modes. In addition, average distance traveled to the fishing site was over 100 miles in the non-boat modes, well over 200 miles in the party/charter boat mode, and under 100 in the private/rental boat mode. About 25% of those intercepted were visitors on vacation, which accounted for the large distances. About 60% of anglers were fishing for a particular species. Of these, most targeted bluefish (22%) and winter and summer flounder (15% and 10% respectively) and, as expected, those with a target species were more avid fishermen. The percentage fishing for a particular species was significantly higher in the private/rental boat mode. The average catch was 6 fish per trip with 4 of these kept. The average catch is higher in the party/charter boat mode as is the percent kept (80%). Most kept fish are used for food, and most not kept are returned to the water alive. Although catching fish is stated as the main reason for fishing on less than 25% of all trips, the level of satisfaction of an angler with the trip is strongly related to the number of fish caught, and particularly to the number kept.

The authors are grateful to Arnold Greenland of IIT Research Institute for reviewing the paper.

REFERENCES

[1] W. Cochran, Sampling Techniques, John Wiley & Sons, 3rd Edition, 1977

[2] D. N. Ghosh, Sampling Design For Marine Recreational Fishing Survey: An Example of a Complemented Survey, Proceedings of the Section on Survey Research Methods, ASA, 1981

[3] Marine Recreational Fishery Statistics Survey, Atlantic and Gulf Coasts, 1979, National Marine Fisheries Serice: Current Fishery Statistics Number 8063

[4] R. L. Hiett, K. A. Chandler, A. K. Reniere, R. Bolstein, Socioeconomic Aspects of Marine Recreational Fishing, National Marine Fisheries Service, May 1983

[5] Nie, Hull, Jenkins, Steinbrenner, and Bent, Statistical Package for the Social Sciences, Second Edition, McGraw Hill, 1975