Indian Journal of Marine Sciences
Vol. 38(4), December 2009, pp. 464-469

On use of post-stratification for estimating the marine fish landings

Mini K G [1*], Kumaran M[2] & Jayasankar J[1]

[1]Fishery Resources Assessment Division, Central Marine Fisheries Research Institute, P.B. No. 1603, Cochin, Kerala 682 018, India

[2]Research Centre of Department of Statistics, Nehru Arts & Science College, Kanhangad, Padnekat P. O. Kerala 671 328, India

[*][E-mail: minikg@rediffmail.com]

Received 26 August 2008 ; revised 20 January 2009

The sampling design adopted by Central Marine Fisheries Research Institute for estimation of marine fish landings in India is a stratified two stage sampling design with stratification being done over space and time. Present study conists a modified design which uses post-stratified estimator to estimate the fish landings in a month in a zone along with its variance. The modified design improved the precision of the estimator as compared to the existing estimator. An empirical study is also carried out to demonstrate the performance of the modified design.

[**Keywords**: Estimation of fish landings, Marine fishery, Post-stratification, Stratified two-stage sampling, Variance estimation]

Introduction

The Central Marine Fisheries Research Institute (CMFRI) developed a stratified multistage sampling design and first put into operation in the State of Kerala in the middle of 1959 and has gradually extended it to other states of the west coast of India. From 1961, the design was introduced along the entire coast of the mainland. During the last five decades the fishery sector has undergone drastic changes, but there were no significant alterations in the basic structure of the sampling design. Hence, evaluation of the sampling design is essential to determine the mode and frequency of data collection keeping in pace with the changing pattern of the fishery. Except for a study[1] there had been no attempt to evaluate the sampling design of CMFRI in terms of the precision of the estimates. The currently used sampling methodology is described in detail by Srinath *et al*[2]. In the existing methodology the gear wise landings were estimated by assuming that all the gears were operating on all days in a month in a zone. But in real situation, it is not so. Apart from that there exists a lot of variation between different types of gears operated; this may result in an over or under estimation amounting to very high discrepancy in the estimate. This problem can be rectified only if information on all the gear types operated in a month at the zone are made available. In the absence of advance information on the gear types operated at the centre, post-stratification can be applied to get more reliable estimates. Post-stratification involves assignment of units into different categories after selection of the sample. Hansen *et al.*[3] were the first to discuss the concept of post stratification. Williams[4] suggested a procedure for getting approximate variance of post-stratified estimator. This aspect has been discussed in case of uni-stage random sampling designs by Murthy[5], Cochran[6], Sukhatme and Sukhatme[7] and others. Post-stratification is well known as a means of increasing the precision of estimates in non-stratified sampling by incorporating additional information about strata weights in the final estimator.

Mehrotra[8] gave a scheme for post-stratification in two stage sampling on the basis of the sampled second stage units. It demonstrates the scheme empirically using a simulated data on area under high yielding varieties of wheat crop in a holding as the character under study. The first stage units being the number of villages and the second stage units the cultivator's holdings growing high yielding variety of wheat in a district. It ascribes that this scheme provides estimates of the character under study according to the strata and improves the precision of the estimate pooled over the strata compared to the conventional non-stratified procedure. The present study consists a modified design which uses the post-stratified estimator given by Mehrotra[8] coupled with the existing sampling design to estimate the fish landings in a month in a zone along with its variance.

Materials and Methods

Outline of the sampling design adopted by CMFRI

The currently followed sampling design is a stratified two stage sampling design with stratification being done over space and time. Each maritime state is divided into several zones on the basis of fishing practices and geographical considerations. A zone is a stratum over space and a calendar month is the stratum over time. There are two types of zones – single centre and multi-centre zones. A group of contiguous landing centres form the multi-centre zone. The number of landing centres may vary from zone to zone. Multi-centre zones are further divided into several strata on the basis of intensity of fishing. The major harbours/centres are classified as single centre zones. In the present study, the focus is made on the estimator of fish landings for the single centre zones. The post stratified strategy can be extended to multi-centre zones provided the auxiliary information on the type and number of gears operated in all the fishing days are known in all the centres in that zone.

A month is divided into 3 groups, each of 10 days. From the first five days of a month, a day is selected at random, and the next 5 consecutive days are automatically selected. From this, three clusters of two consecutive days are formed. For example, for a given zone, in a given month, from the five days if the date (day) selected at random is 4, then the clusters formed from the first 10 day group are (4, 5), (6, 7) and (8, 9). In the remaining ten day groups, the clusters are systematically selected with an interval of 10 days. For example, in the above case, the clusters of days for observation in the remaining groups are (14, 15), (16, 17), (18, 19); (24, 25), (26, 27) and (28, 29). Normally, in a month 9 clusters of two days each can be obtained. A combination of landing centre and day known as landing centre day constitute primary stage unit. The landing centre day is the 24 hour duration from the mid day of the first day to the mid day of the following day in the landing centre.

A landing centre day has been divided into 3 periods as given below.

Period	Duration
Period 1	1200 to 1800 hours on 1st day
Period 2	0600 to 1200 hours on 2nd day
Period 3	1800 hours to next morning 0600 hours

One field staff is usually provided to each zone. A field staff starts data collection from period 1 on each selected day. The enumerator will be present through out the periods 1 and 2 at the centre. The data on landings during period 3 (night landings) is usually collected from the landing centre by enquiry on the following day of period 1. The sum of the observations on the 3 periods contributes the data for the day.

The fishing boat that land on a landing centre day forms the second stage units. The field staff after reaching the landing centre first gathers information on the probable number of boats which are expected to land at the centre on that day. If the number of boats to be landed is large, it may not be practicable to record the catches of all boats landed during an observation period. A sampling of the boats then becomes essential. When the total number of boats landed is 15 or less, the landings from all the boats are observed for catch and other particulars. When it exceeds 15, the following procedure[9] is followed.

Number of boats landed	Fraction to be observed
Less than or equal to 15	100%
Between 16 and 19	First 10 and the balance 50%
Between 20 and 29	1 in 2
Between 30 and 39	1 in 3
Between 40 and 49	1 in 4
Between 50 and 59	1 in 5 and so on

In case the number of boats landed is very high, the field staff may not be able to stick to the above condition. In such situations, the observations are restricted to a maximum of 15 boats selected as above. From the boats, the catches are normally removed in baskets of standard volume. The weight of fish contained in these baskets being known, the total weight of the fish in each boat under observation has been obtained. The data on species-wise landings, effort, type of crafts and gears operated and nature of fishing ground are collected. Based on the information from the selected fishing units, the total landings for the day are estimated. From the estimates of day landings, monthly estimates for the zone are worked out. For the estimation purpose, both the first stage and second stage units are assumed to be selected with simple random sampling without replacement.

Let N be the number of days (fishing days) in a month and n be the number of selected days out of the N days.

Let M_{ikp} be the total number of boats landed during p^{th} period of observation of gear type k on the i^{th} selected day. ($k = 1,2,...,T_i$ (T_i is the number of distinct gear types landed on the i^{th} selected day); $i = 1,2,...,N$; $p = 1,2,3$). Let m_{ikp} be the number of selected boats during the p^{th} period of observation of gear type k on the i^{th} landing centre day and $m_{ik.}$ denote the total number of boats of the k^{th} type gear sampled on i^{th} day.

Let y_{ikpl} be the quantity of fish landed by the l^{th} selected boat during p^{th} period of observation of k^{th} gear type on i^{th} selected day. ($l = 1,2,...,m_{ikp}$)

Let $\overline{y}_{ikp}$ denote average quantity of fish landed during p^{th} period of observation by the k^{th} type gear on i^{th} selected day, which is given by

$$\overline{y}_{ikp} = \frac{1}{m_{ikp}} \sum_{l=1}^{m_{ikp}} y_{ikpl} \qquad \ldots (1)$$

Let $\hat{Y}_{ikp}$ denote the estimated total landings during p^{th} period of observation by k^{th} type gear on i^{th} selected day, which is given by

$$\hat{Y}_{ikp} = M_{ikp}\ \overline{y}_{ikp} \qquad \ldots (2)$$

Let $\hat{Y}_{ik.}$ be the estimated total landings by k^{th} gear type on i^{th} selected day, then

$$\hat{Y}_{ik.} = \sum_{p=1}^{3} \hat{Y}_{ikp} \qquad \ldots (3)$$

The estimated total landings ($\hat{Y}_{i.}$) for the i^{th} day is obtained as

$$\hat{Y}_{i..} = \sum_{k=1}^{T_i} \hat{Y}_{ik.} \qquad \ldots (4)$$

The estimated total landings ($\hat{Y}_{.k.}$) by the k^{th} gear type for the month is obtained as

$$\hat{Y}_{.k.} = N\ \hat{\hat{Y}}_{.k.} \qquad \ldots (5)$$

Where $\hat{\hat{Y}}_{.k.} = \frac{1}{n}\sum_{i=1}^{n} \hat{Y}_{ik.}$ is the average landings by the k^{th} gear type per day.

Since the gear types vary with respect to days, the number of distinct gears *T* over the days in the month may be taken for estimating the total landings. Hence, $\hat{Y}$, the estimated total landings for the month over all gear types is given by

$$\hat{Y} = \sum_{k=1}^{T} \hat{Y}_{.k.} \qquad \ldots (6)$$

Assuming that the variance between gears and between boats of the same gear are negligible[8] within a selected day, the estimate of variance of the total landings is given by

$$\hat{V}(\hat{Y}) = N^2 \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{n-1} \sum_{i=1}^{n} \left(\hat{Y}_{i..} - \hat{\hat{Y}}_{...}\right)^2 \qquad \ldots (7)$$

where $\hat{Y}_{i..}$ is total landings given by equation (4) and $\hat{\hat{Y}}_{...}$ is the average landings per day. The standard error of the estimate can be found out from the above formula.

The Modified Design

This design is intended only for single centre zones. One important modification of the modified design is regarding the 24 hr duration of the day as a single unit for the purpose of estimation instead of regarding it as divided into three periods. The existing procedure of selecting the boats for observation as described earlier may be substituted in the modified design. The number of the boats landed to be recorded continuously through out the day. At the end of the day, the number of boats landed and the catches recorded are post-stratified according to the gear types. The number of distinct gears, *T*, may be taken as the number of post-strata. The resulting design can be regarded as a two stage random sampling with post-stratification at the second stage with first stage unit as landing centre day and second stage unit as boats of specific gear type. To analyse the data the existing procedure coupled with the procedure for post-stratification by Mehrotra[6] had been followed.

Out of the N fishing days in a month, n are selected at random for observation. Let the observed number of days containing at least one fishing boat belonging to j^{th} gear be denoted by $n_{(j)}$, ($0 < n_{(j)} \leq n$) and $m_{i(j)}$ denote the sampled number of fishing boats of the j^{th} gear landed on the i^{th} day. Similarly $N_{(j)}$ denote the total number of days j^{th} gear landed and $M_{i(j)}$, the total number fishing boats of the j^{th} gear observed.

Analogous to the existing estimator (5), an unbiased estimator of the total fish landings by the j^{th} type gear is given by

$$_{post}\hat{Y}_j = \frac{N_{(j)}}{n_{(j)}}\sum_{i=1}^{n_{(j)}}\frac{M_{i(j)}}{m_{i(j)}}\sum_{k=1}^{m_{i(j)}} y_{ik(j)} \quad \ldots (8)$$

where $y_{ik(j)}$ is the quantity of fish landed by the k^{th} fishing boat of the j^{th} type on the i^{th} day. (Note that in (8) there is no period-wise summation as in (5) since the day is treated as a single unit.)

The total landings for the zone in a month is given by

$$_{post}\hat{Y} = \sum_{j=1}^{T}\frac{N_{(j)}}{n_{(j)}}\sum_{i=1}^{n_{(j)}}\frac{M_{i(j)}}{m_{i(j)}}\sum_{k=1}^{m_{i(j)}} y_{ik(j)} \quad \ldots (9)$$

The variance of the estimator $_{post}\hat{Y}$ is given by, Mehrotra[6],

$$V\left({}_{post}\hat{Y}\right) = \sum_{j=1}^{T} N^2\left(\frac{1}{n}-\frac{1}{N}\right) w_{(j)} S_{(j)}^2$$

$$+\left[\frac{N^2}{n^2}\sum_{j=1}^{T}\left(1-w_{(j)}\right)S_{(j)}^2\right] N^2 \sum_{j\neq j}^{T} w_{(jj')}$$

$$+\left(\left(1+\frac{1-w_{(j)}}{nw_{(j)}}+\frac{1-w_{(j')}}{nw_{(j')}}+\frac{1}{n}\right)\left(\frac{1}{n}-\frac{1}{N}\right)-\left(\frac{1+w_{(jj')}}{N_{(j)}}\right)\right)$$

$$+S_{(jj')}\, N^2\sum_{j=1}^{T}\frac{1}{nN}$$

$$\left(1+\frac{1-w_{(j)}}{nw_{(j)}}\right)\sum_{i=1}^{N_{(j)}}\left(M_i^2\left(w'_{i(j)}\right)\left(\frac{1}{m_i}-\frac{1}{M_i}\right)+\frac{1-w'_{i(j)}}{m_i^2}\right)S_{i(j)}^2 \quad \ldots (10)$$

Where

$$w_{(j)} = \frac{N_{(j)}}{N},\ w_{(j')} = \frac{N_{(j')}}{N},\ w_{(jj')} = \frac{N_{(jj')}}{N},\ w'_{i(j)} = \frac{M_{i(j)}}{M_i}$$

where $N_{(jj')}$ is the total number of days having fishing boats belonging to gear types j^{th} and $j'^{\,th}$ landed, $n_{(jj')}$ the corresponding number in the sample.

$$S_{(j)}^2 = \frac{1}{N_{(j)}-1}\sum_{i=1}^{N_{(j)}}\left[Y_{i(j)}-\overline{Y}_{(j)}\right]^2 \quad \ldots (11)$$

$$S_{(jj')} = \frac{1}{N_{(jj')}-1}\sum_{i=1}^{N_{(jj')}}\left[Y_{i(j)}-\overline{Y}_{(j)}\right]\left[Y_{i(j')}-\overline{Y}_{(j')}\right] \quad \ldots (12)$$

$$S_{i(j)}^2 = \frac{1}{M_{i(j)}-1}\sum_{j=1}^{M_{i(j)}}\left[Y_{ik(j)}-\overline{Y}_{i(j)}\right]^2 \quad \ldots (13)$$

It can be noted that equation (10) is made up of four components. The first term appears to be the variance of a stratified sample taken with proportional allocation at the first stage, the second represents the adjustment at the first stage due to post stratification of the sampled first stage units (days) and the last two terms represent contribution to the variance on account of the stratification of the second stage units (fishing boats) and the adjustment at the second stage due to post stratification of the second stage units.

An unbiased estimate of $V\left({}_{post}\hat{Y}\right)$ is given by

$$\hat{V}\left({}_{post}\hat{Y}\right) = \sum_{j=1}^{T} N^2\left(\frac{1}{n}-\frac{1}{N}\right) w_{(j)} s_{(j)}^2$$

$$+\sum_{j\neq j}^{T} N_{(j)}N_{(j')}\left(\frac{n_{(jj')}}{n_{(j)}n_{(j')}}-\frac{n_{(jj')}^2}{n_{(j)}n_{(j')}N_{(jj')}}\right)s_{(jj')}$$

$$+\sum_{j=1}^{T}\frac{N_{(j)}}{n_{(j)}}\sum_{i=1}^{n_{(j)}} M_{i(j)}^2\left(\frac{1}{m_{i(j)}}-\frac{1}{M_{i(j)}}\right)s_{i(j)}^2 \quad \ldots (14)$$

Where

$$s_{(j)}^2 = \frac{1}{n_{(j)}-1}\sum_{i=1}^{n_{(j)}}\left(y_{i(j)}-\overline{y}_{(j)}\right)^2 \quad \ldots (15)$$

$$s_{(jj')} = \frac{1}{n_{(jj')}-1}\sum_{i=1}^{n_{(jj')}}\left(y_{i(j)}-\bar{y}_{(j)}\right)\left(y_{i(j')}-\bar{y}_{(j')}\right) \quad \ldots (16)$$

$$s_{i(j)}^{2} = \frac{1}{m_{i(j)}-1}\sum_{k=1}^{m_{i(j)}}\left(y_{ik(j)}-\bar{y}_{i(j)}\right)^{2} \quad \ldots (17)$$

are the unbiased estimators of the respective population mean squares.

Equation (9) gives the estimator of the total fish landings/month in the zone and equation (14) the estimator of its variance based on the new design. Note that the design as proposed to a single centre zone allows enough scope to ensure any desired sampling fraction by simply altering the first stage sample size n. Again due to the post-stratification, the major source of variation due to gears are also well accounted. The procedure can also provide gear-wise estimators as given by equation (8).

An Empirical Illustration

The marine fish landings data at Cochin fisheries harbour during the year 2004 is used for the illustration. The available fish landings data was collected by using the existing two stage design. The number of first stage units are 9 landing centres days in a month and the second stage units varies over the selected landing centre days. The important gears operating at Cochin fisheries harbour are mechanized trawl nets, mechanized gillnets, mechanized hooks & lines, purse seines and motorized ring seines. The estimate of marine fish landings for each month during the year was found out both by the existing procedure and also by the new modified procedure. In the modified estimation procedure, post stratification was done according to the gear used for fishing. One of the greatest practical limitations to the use of post-stratification is the need to know the total number of units in each strata. Since the existing data is in terms of crafts which use multiple gears, explicit gear wise data are not available. To overcome this difficulty we construct a population for the number of gears landed. For this, we proceed as follows. Firstly, the number of each type of gear landed for a month was taken from the sample collected. Further, the number of different crafts landed on each day in the harbour was collected from the register maintained by the Cochin Port Trust. If the craft-wise estimates of landings were of interest, then the above data can be directly used for the estimation purpose. In this illustration, we focus on the gear-wise estimates. Hence, based on the data on the number of crafts landed from Cochin Port Trust and the data collected by CMFRI, the proportion of each gear type was made for a month. Then, the number of each gear type landed was simulated by assuming that it will follow a multinomial distribution. This information was used for the estimation procedure. (Comparison of the estimates of marine fish landings at cochin fisheries harbour obtained through the existing CMFRI adopted sampling design and the modified design is ascribed in the **Table 1**).

Conclusion

The monthly estimates of marine fish landings at Cochin fisheries harbour and its variance were

Table 1—Comparison of the estimates of marine fish landings at cochin fisheries harbour obtained through the existing CMFRI adopted sampling design and the modified design

	CMFRI adopted Design			Modified Design		
Month*	Estimate	Variance	CV	Estimate	Variance	CV
Jan	362	4112	18	380	2104	12
Feb	479	11301	22	575	7287	15
Mar	883	100427	36	1053	59428	23
Apr	531	43730	39	780	67560	33
Jul	551	63599	46	495	32169	36
Aug	4895	120478	7	4975	83213	6
Sep	3460	1754069	38	4646	1671017	28
Oct	2347	235540	21	3054	210272	15
Nov	631	47527	35	637	37117	30
Dec	457	3546	13	474	2710	11

CV-Coefficient of Variation

*Data on number of boats landed was not available for the month of May and June.

computed using the new sampling design. These estimates were compared with the estimates obtained by the existing sampling design. The table lists the estimated marine fish landings, the variance and the corresponding coefficient of variation for the existing sampling design and the new design. The coefficient of variation based on the new design is smaller than that of the existing design for all months. In some months the reduction in the coefficient of variation was even greater than 10%. The numerical results clearly indicate that the post-stratified estimator significantly outperformed and it is more efficient than the currently used estimator. Hence, the modified design and the corresponding estimators are suggested for use whenever the auxiliary information on the type and number of gears operated in a day are known in single centre zones.

Acknowledgement

The authors are grateful to the Director, CMFRI for extending necessary facilities for the study. The first author thank Ministry of Statistics and Programme Implementation, Government of India for the award of doctoral fellowship during the study.

References

1 Kutty M K, Nair A K K & Qasim S Z, An evaluation of the sampling design adopted by the CMFRI for estimating marine fish production in India, *Indian J. Fish*. 22(1973) 16-34.
2 Sukhatme P V, Panse V G & Sastry K V R, Sampling techniques for estimating catch of sea fish in India, *Biometrics*, 14(1958) 78-96.
3 Hansen M H, Hurwitz W N & Madow W G, Sample Survey Methods and Theory, Vol I & II, (John Wiley & Sons, New York) 1953.
4 Williams W H, The variance of an estimator with post stratified weighting, *J. Amer. Stat. Assoc*. 57(1962) 622-627.
5 Murthy M N, Sampling Theory and Methods, (Statistical Publishing Society, Calcutta, India) 1967.
6 Cochran W G, Sampling Techniques, (John Wiley & Sons, New York) 1977.
7 Sukhatme P V & Sukhatme B V, *Sampling Theory of Surveys with Applications*, (J. Ind. Soc. Agril. Stat., New Delhi) 1976.
8 Mehrotra P C, A scheme for post stratification in two stage sampling, *Jour. Indian. Soc. Agri. Stat*., 45(1993), 163-170.
9 Alagaraja K, Simple methods for estimates of parameter for assessing exploited fish stocks, *Indian J. Fish*., 31(1984) 177-208.