A STUDY ON EVALUATION OF CMFRI SAMPLING DESIGN FOR ESTIMATION OF MARINE FISHERIES RESOURCES IN KERALA

THESIS SUBMITTED TO THE KANNUR UNIVERSITY
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
STATISTICS

BY

MINI K. G.

UNDER THE SUPERVISION OF
DR. M. KUMARAN

RESEARCH CENTRE
DEPARTMENT OF STATISTICS
NEHRU ARTS & SCIENCE COLLEGE, KANHANGAD
KERALA - 671 328

MAY 2008

Dr. M. Kumaran
Reader
Research Centre
Department of Statistics
Nehru Arts & Science College
Kanhangad
Kerala - 671 328

CERTIFICATE

This is to certify that the thesis "**A Study on Evaluation of CMFRI Sampling Design for estimation of Marine Fisheries Resources in Kerala**" submitted to the University of Kannur for the award of degree of **Doctor of Philosophy** in **Statistics** is a record of original research work carried out by Ms. Mini K.G during the period of her study at the Research Centre of Department of Statistics, Nehru Arts and Science College, Kanhangad, Kerala under my guidance and supervision and the thesis has not so far been formed the basis of award of any Degree / Diploma / Associateship / Fellowship or any other similar title to any candidate of any University or Institute.

Kanhangad
May 23, 2008

(Dr. M. Kumaran)

Dr. M. KUMARAN,
Reader and Research Guide.
Research Centre,
Dept. of Statitics
Nehru Arts & Science C
Kanhangad

DECLARATION

I hereby declare that the matter embodied in this Thesis is the result of investigations carried out by me in the Research Department of Statistics, Nehru Arts & Science College, Kanhangad under the supervision and guidance of Dr. M. Kumaran, Reader, Research Centre, Department of Statistics, Nehru Arts & Science College, Kanhangad and it has not been submitted for award of any Degree/Diploma/Associateship/Fellowship of any other University or Institute.

Kanhangad

May 23, 2008

(Mini K. G.)

Acknowledgements

First and foremost I would like to thank the Almighty for making my academic aspirations and accomplishments a reality.

My sincere and utmost gratitude goes to my supervisor Dr. M. Kumaran, Reader, Research Centre, Department of Statistics, Nehru Arts & Science College, Kanhangad for his invaluable support and patient guidance. His unparalleled gestures to kindness made me at most ease during the period of this study. Indeed his relentless efforts have given this thesis a shape.

I am highly grateful to Dr. Mohan Joseph Modayil, former Director, C.M.F.R.I. for granting me study leave and the present Director, Dr. N. G. K. Pillai for their support and encouragement for the completion of this work.

I am also indebted to Ministry of Statistics and Programme Implementation for offering me the Doctoral fellowship to pursue this study.

I owe a great deal to Dr. M. Srinath, Principal Scientist, N.R.C. for Women, Orissa who has been a source of inspiration and encouragement. I avail this opportunity to express my profound gratitude Dr. J. Jaysankar, Senior Scientist, C.M.F.R.I. for evincing keen interest in my work and suggesting stimulating ideas during this study. I am extremely grateful to my friend Dr. Somy Kuriakose, Senior Scientist, C.M.F.R.I. for helpful advice in every phase of my study. I am thankful to her continuing support in my academic development.

My sincere and heartfelt thanks are also due to Dr. A. K. Srivastava, Former Joint Director, I.A.S.R.I whose inspiring suggestions benefited me a lot during the course of the study. I am extremely grateful to Dr. N. Kurup, Shri. Balan, Dr. G. K. Jha, Dr. Radhakrishnan Nair, Dr. Sathyanandan, Dr. Ramasubramanian, Dr. Dandapani and Smt. Lasitha for their invaluable suggestions which proved vital for this work.

I would also like to express my profound thanks to Dr. Raju, Reader, Department of Statistics, Calicut University for providing the right support in finding my research guide.

My special thanks goes to my well wishers Smt. Valsala, Dr. Tara Thomas, Dr. Sethi, I.A.S.R.I. and his family, Dr. Femina, C.I.F.T, Cochin and her family and Shri. Vijayan Thidunnayil and family for their co-operation and whole-hearted support.

I am really very happy to extend my heartfelt appreciation to Dr. Fred and Dr. George Mammen of St. Pious College, Rajapuram and their family for the exceptional generosity, care and support they have given, for which, there will be always a dearth of words to describe.

I express my deep gratitude to Smt. Beena V. T. and Smt. Mercey Joseph for their constant guidance, enthusiastic encouragement and insightful suggestions.

My sincere thanks are also due to fellow researchers, Shri. Vijayan and Smt. Pushpaja and Smt. Geetha of Department of Statistics for their generous assistance in the research proceedings. I am also thankful to the Librarian and other staff of Nehru Arts & Science College, Kanhangad, C.M.F.R.I, Cochin, I.A.S.R.I, New Delhi and Kerala University, Thiruvananthapuram for their valuable help in procurement of relevant literature.

The paintaking efforts of Smt. Beena M. R. and Smt. Sindhu Augustine in the collation of data need special mention and a word of appreciation. My thanks are also due to Smt. Ammini, Smt. Ramani, Shri. Prasad, Shri. Narayanankutty, Shri. Pavithran and all my other colleagues in the Fishery Resources Assessment Division for their help and encouragement.

It gives me great pleasure to express my heartfelt gratitude to my friends to Vipin Kumar, Thanooja, Anitha, Nishitha, Harish, Sandeep and Joji for their help during the course of this study. My special thanks also goes to my friends at Kanhangad, Sandhya, Nirmal, Sonia, Sruthi and Reshma and all others who helped me in course of the study.

I am grateful to my family for their unfailing love and support. I would like to thank my beloved father Shri. Gopalan, my two brothers, sister-in-laws and relatives for their constant encouragement. Although my mother Smt. Sulochana left this world, her unfailing love and expectation still greatly inspires me in pursuing more knowledge.

I want to thank from the bottom of my heart for the continuous encouragement and indescribable support of my beloved father-in-law, Shri. Bose and mother-in-law, Smt. Sobhana. I specially thank my mother-in-law who managed all the family affairs herself making it possible for me to concentrate on my studies without which this research work would not have been possible. I dearly thank my brother-in-law for his help and support.

I must acknowledge with great affection the whole-hearted love and support of Smt. Leela who has taken utmost care of my son during the period of my study.

I express my entire affection for my dear husband, Shri. Anish, the most important person in my life, who has always been a source of inspiration and pillar of strength. Without his immense understanding and support, this study would not have been completed. My deep appreciation to him is inexpressible in any language.

Finally and most importantly, I would like to dedicate this work to my beloved children, Akshadha and Adinath, who could not avail the required care at this tender age. The maturity shown in accepting my absence and the silent support of these tiny tots was the real motivation for completing the study within the specified time.

CONTENTS

I		**Introduction and Review of Literature**	**1-29**
	1.1	Introduction	1
	1.2	Central Marine Fisheries Research Institute	2
	1.3	Development of CMFRI Sampling Methodology	7
	1.4	Marine Fisheries of Kerala	14
	1.5	Objectives of the Study	19
	1.6	Review of Literature	21
	1.7	Plan of the Thesis	27
II		**CMFRI Sampling Design in Kerala – An Evaluation**	**30-71**
	2.1	Introduction	30
	2.2	CMFRI Sampling Design	30
	2.2.1	Selection of Primary Stage Units	34
	2.2.2	Period of Observation	35
	2.2.3	Selection of Second Stage Units and Recording of Fish Landings	35
	2.2.4	Administration of the Survey	36
	2.2.5	Analysis of Data	38
	2.3	Zonal Sampling Fraction and Estimates of Landings	49
	2.4	Limitation of the Existing Design	63
	2.5	Remedial Measures	70
III		**Marine Fish Landings in Kerala – Fluctuations and Trends**	**72-98**
	3.1	Introduction	72
	3.2	Changes in the Fisher Folk Population, Crafts and Gears	72
	3.3	Trends in the Total Marine Fish Landings	75
	3.4	Zone Level Trends	81
	3.5	Comparison of Period-wise Landings at Single Centre Zones	93
	3.6	Conclusion	98

IV		**Estimation of Variance Components by the Method of Nested Analysis**	**99-133**
	4.1	Introduction	99
	4.2	Analysis of Hiearchical or Nested Data	100
	4.3	Nested Analysis of Fishery Data	116
	4.4	Conclusion	131
V		**Estimation of Fish Landings - Post-Stratified Designs**	**134-155**
	5.1	Introduction	134
	5.2	Post-Stratified Design for Single Centre Zones	135
	5.3	Post-stratified Design for Multi-Centre Zones	142
	5.4	Conclusion	154
VI		**Estimation of Fish Landings – PPS Sampling Design**	**156-179**
	6.1	Introduction	156
	6.2	Two Stage PPS Design	158
	6.2.1	PPS Estimator of Landings per Craft	159
	6.2.2	PPS Estimator of Landings per Day	163
	6.2.3	Sampling Fraction	167
	6.2.4	Estimation of PPS Selection Probabaility	168
	6.3	Estimation of Optimum Sample Size	169
	6.3.1	Optimum Sample Size for Multi-Centre Zone in the Existing Design	171
	6.3.2	Optimum Sample Size for Two Stage PPS Design	173
	6.4	Conclusion	178
		Concluding Remarks and Recommendations	**180-185**
		References	**186-194**

List of Tables

Table No.	Title	Page No.
2.1(a)	Distribution of the landing centres over the coastline districts	31
2.1(b)	Zonal distribution of landing centres	33
2.2	Fish landings data as a two-way table with *n* days and T_l gears during p^{th} period of observation	40
2.3	Adjusted data during the p^{th} period of observation	41
2.4	(1) to (15) Zonal sampling fractions during 2004-06	50-62
2.5	Estimated sample sizes for specified margin of error for the year 2006	65
3.1	Fisher folk population 1980-2005	73
3.2	Fisher folk population involved in fishing as per census 1980 and 2005	73
3.3	Number of craft in the fishery during 1998-2005	74
3.4	Results of t-test for comparison of landings during afternoon and forenoon – 2006	94
4.1	Analysis of variance table for the *s*-level nested data	112
4.2	Analysis of variance for the marine fish landings data for the zone K1 during 2006	121-122
4.3	(1) to (15) Zonal monthly estimate of average landings (in kg.) and the standard error for the year 2006	123-130
4.4	Zone-Month exhibiting more than 10% variation due to days and boats	130-131
5.1	Comparison of existing design and design I	141

List of Figures

Figure No.	Title	Page No.
2.1	Geographical distribution of the fishing zones in Kerala	32
3.1	(a) to (e) Trend in total marine fish landings in Kerala during 1961-2006	77-78
3.2	Pelagic and demersal fish landings during 1961-2006	79
3.3	Contribution of different sectors to the marine fish landings	80
3.4	(a) to (f) Trend in the landings at zone K1 during 2002-06	82-83
3.5	(a) to (d) Trend in the fishing boats operated at zone K1 during 2002-06	83-84
3.6	(a), (b) Trend in the landing per boat at zone K1 during 2002-06	84
3.7	(a) to (i) Trend in the landings at zone K6 during 2002-06	85-86
3.8	(a) to (f) Trend in the fishing boats operated at zone K6 during 2002-06	86-87
3.9	(a) to(c) Trend in the landing per boat at zone K6 during 2002-06	87
3.10	(a) to (i) Trend in the landings at zone K8 during 2002-06	88-89
3.11	(a) to (f) Trend in the fishing boats operated at zone K8 during 2002-06	89
3.12	(a) to(c) Trend in the landing per boat at zone K8 during 2002-06	90
3.13	(a) to (f) Trends in marine fish landings at Neendakara during 2002-06	91
3.14	(a) to (d) Trends in number of boats landed at Neendakara during 2002-06	92
3.15	(a), (b) Trends in landings per boat at Neendakara during 2002-06	92
3.16	(a), (b) Comparison of boats landed and average landings between afternoon and forenoon at Neendakara 2005	95
3.17	(a), (b) Comparison of boats landed and average landings between afternoon and forenoon at Neendakara 2006	95

3.18	(a), (b) Comparison of boats landed and average landings between afternoon and forenoon at Sakthikulangara 2005	96
3.19	(a), (b) Comparison of boats landed and average landings between afternoon and forenoon at Sakthikulangara 2006	96
3.20	(a), (b) Comparison of boats landed and average landings between afternoon and forenoon at Cochin Fisheries Harbour 2005	97
3.21	(a), (b) Comparison of boats landed and average landings between afternoon and forenoon at Cochin Fisheries Harbour 2006	97
4.1	Nested structure of marine fish landings data	118

CHAPTER I

Introduction and Review of Literature

1.1 Introduction

Fishing has been a traditional occupation of a section of people all over the world from time immemorial. In India, the fisheries sector contributes significantly towards strengthening nutritional security, income, employment, foreign exchange earnings and livelihood opportunities. These facts established the fisheries sector as an important enterprise of Indian economy. During the last six decades, Indian fisheries had made tremendous progress, with the annual fish production increasing from 0.75 million tonnes in 1950 to 6.4 million tonnes in 2006, indicating over eightfold increase during the period. As per the latest available data, marine fishery sector earns a foreign exchange of Rs. 8000 crores annually through seafood export.

The fish production in the sea is depended upon the productivity of the sea, the availability of fish at a given point of time, the fishing effort expended, accessibility and vulnerability of the resources and a number of other factors. Man has a lot of control on land-based resources whereas in marine fishery resources the only opportunity for man to intervene is through management of the capture process. In the case of agricultural crops, production can be increased by means of high quality seeds, fertilizers, irrigation, pest management and so on. Unlike the land-based resources, marine fisheries resources are invisible, frequently migrating and easily affected by the changes in the sea. These characteristics make it unique and complex and hence difficult to monitor, manage and intervene. This uniqueness of marine fisheries makes it a challenging task for scientists to explore the dynamics of the fishery and the fishery managers to make management interventions. For interventions aimed at any developmental or management agenda in this sector constant and continuous monitoring of the resources is essential.

1.2 Central Marine Fisheries Research Institute

Recognising the importance of the fisheries sector to the state's economy, the need to establish a Fisheries Research Station in the country had become very strong. Accordingly, a Fisheries Research Station was established on 3rd February 1947 in the University of Madras. Following this, several fisheries research stations were established in different states of the country under the Ministry of Food and Agriculture. The Fisheries Research Station at Madras was shifted to Mandapam Camp, Tamil Nadu in 1949 and became the head quarters of the Research Stations. Later on this station was developed into a full-fledged Marine Fisheries Research Institute and renamed as Central Marine Fisheries Research Institute (CMFRI) in 1962. The Institute is mandated to carryout research and developmental activities in marine fisheries. In October 1967, the management and administrative control of the Institute was transferred from the Ministry of Food and Agriculture to the Indian Council of Agricultural Research (ICAR). The headquarters of the Institute was shifted to Cochin, Kerala in 1971. The CMFRI contributed greatly to the understanding of fishery biology and fishery oceanography enabling the rational exploitation of several fish stocks. The institute is mandated

(i) to monitor the exploited and assess the under-exploited marine fisheries resources of the Exclusive Economic Zone,

(ii) to understand the fluctuations in abundance of marine fisheries resources in relation to changes in the environment,

(iii) to develop suitable mariculture technologies for finfish, shellfish and other culturable organisms in open seas to supplement capture fishery production,

(iv) to act as a repository of information on marine fishery resources with a systematic database,

(v) to conduct transfer of technology, post graduate and specialized training, education and extension programmes and

(vi) to provide consultancy services.

To carry out these tasks effectively, the Institute has established Regional Centres at Mandapam Camp, Veraval and Visakhapatnam, Research Centres at Minicoy, Mumbai, Karwar, Mangalore, Kozhikode, Vizhinjam, Tuticorin, Chennai and Kakinada and 28 Field Centres all along the coastal line of the country. The entire activity is coordinated by the headquarters at Cochin. The CMFRI has, over the years, built up laboratory and field facilities at all its centres for carrying out research programmes and has been upgrading the same to meet the changing needs and additional requirements. The multidisciplinary researches in capture and culture fisheries are conducted under following ten divisions:

- Fisheries Resources Assessment
- Pelagic Fisheries
- Demersal Fisheries
- Crustacean Fisheries
- Molluscan Fisheries
- Fishery Environment Management
- Marine Biotechnology
- Socio-Economic Evaluation and Technology Transfer
- Mariculture
- Marine Biodiversity

Inter-divisional and inter-institutional programmes were carried out for greater utililisation of expertise and facilities. Besides, the Institute takes up research projects on important and priority areas funded by outside agencies in the country and abroad, and offers consultancy services to the clients from government organisations as well as industry.

The marine living resources are dynamic and renewable and hence regular assessment and monitoring of factors like their resource size, dynamics, exploitation rates and replenishment capacities are essential.

The management of marine living resources requires time series data on all these factors. The major mandate of CMFRI is to monitor and assess the exploited marine fishery resources and render policy support to the Union and State Governments. The data on catch, effort and biological aspects are the essential requirements for assessing the exploited stock. In India, we have a multi-species, multi-locational, multi-gear, seasonal fishery which is being exploited through an open access regime without any serious management interventions. Fish landings take place all along the coast line in all seasons during day and night. In such a complex situation, the collection of landing statistics becomes a formidable task. The cost, operational difficulties and non-sampling errors of a continuous survey covering all landing centres would be of very high magnitude. Hence, a scientifically planned sampling strategy is the only answer to enable estimation of landings by the large number of mechanised and non-mechanised units operating in the coastal belt.

Soon after its inception in 1947, attempts were made by the Fishery Research Station to evolve scientific methods of collecting marine fish catch statistics. In the beginning, not much information was available on the marine fishing villages, landing centres, fishing crafts and gears which could form a frame for developing sampling plans. Also, fishing practices differed from region to region and also from season to season. Keeping this in view, CMFRI along with the ICAR conducted a series of pilot surveys to collect such information as was required for formulating a sampling design. The limited resources at the disposal of the Institute were another constraint in conducting large-scale surveys. However, a stratified multistage random sampling design was developed by the Institute for estimating the exploited fish stocks which became a landmark and was adopted by the Food and Agricultural Organisation of the United Nations for use in other countries. The stratified multistage sampling design was first put into operation in the State of Kerala in the middle of 1959 and was gradually extended to other states of the west coast of India. From 1961, the design was introduced along the entire coast of the mainland. The sampling scheme employed for the estimation of marine fish landings was basically the same but it varied to some extent in details from region to region in view of the varying field conditions. In tune with the fast changing marine fisheries scenario, the scope, the structure and administration of the resource data collection was periodically modified.

The Fisheries Resources Assessment Division of the Institute is primarily responsible for the fulfilment of the Institute's mandate on the monitoring and assessment of the exploited marine fishery resources in the Indian Exclusive Economic Zone. The development of methodologies on data collection for fishery monitoring and assessment were done through a continuing research project titled "Assessment of Exploited Marine Fishery Resources". The project aims to arrive at an estimate of marine fish landings and fishing effort in different regions of the country with species-wise and gear-wise break up of the exploited resources. It also envisages maintaining and updating the database on the Marine Information System existing at the Institute. At present a stratified two stage sampling design was employed to collect and estimate the landings of the exploited marine fishery resources. The planning, execution and co-ordination of field work, processing of data and updating database, developing suitable formats for storage and retrieval are done by the division.

1.3 Development of CMFRI Sampling Methodology

Pilot surveys

The first attempt to build up a planned survey for the estimation of fish catch on an all-India basis was made by the CMFRI. In the pilot survey conducted in 1948-49, village-wise data were collected on the area exploited, the number of persons engaged in marine fishing, the number of various types of fishing boats and nets, fishing seasons, type of fish caught and the number of landing centres. This brought forth a complete picture of fishing activities. Afterwards fisheries data were collected on regular basis from 1950 onwards by dividing the entire coastline into twelve homogenous survey zones. With the availability of more funds and additional staff, the survey zones were increased to cover more landing centres. Between 1950 and 1956, the ICAR also initiated a number of pilot surveys with various designs in different regions of the country with a view to evolve the most suitable sampling design for the estimation of fish landings in the country.

The pilot surveys and their results have influenced a great deal in moulding the sampling design currently used by CMFRI.

In 1950-51 a pilot survey was undertaken in Malabar coast over a coastline of hundred miles. As in any sampling problem, the first efforts were to define the population to be sampled and an appropriate sampling unit. A fixed number of fishing boats were selected from each village in the coast which were kept under observation over time for estimating:

(i) the percentage number of times they went out for fishing and

(ii) the average catch per boat on the basis of sub samples of these boats.

But in a village the number of boats was a highly variable factor; further boats of a village did not often land in the same village. So the practice of selecting village was abandoned. In its place it was found easier to consider the boats where they had landed or first became available for recording. Finally it was agreed that the fishery data may be based on the number of distinct landing places each of which can be considered as a sampling unit. The number of boats landed in the landing centre over a period was determined and the average catch per boat was estimated on the basis of a sub sample of boats landed. The total catch for the period was then estimated as the product of the average catch per boat and the total number of boats landed.

A group of continuous landing centres formed a stratum in space and weeks provided time strata. A landing centre was the primary sampling unit

(PSU). Twenty-minute time intervals were the secondary units (SSU) and an operating fishing boat was the tertiary unit (TSU). The SSU was the ultimate sampling unit for estimating the count and the TSU was the ultimate sampling unit for observing catch. The observations on count and catch were made in distinct intervals. Thus a three stage stratified sampling design was evolved for the collection of marine fishery data.

In 1953-54, another survey was carried out on Malabar coast. The coverage was also extended to include 60-mile coastal part of the southern half of the South Canara district (Karnataka State), as it is geographically contiguous and has similar fishing practices and conditions to those on the Malabar coast. Here the design was one of the 3-stage stratified sampling for recording data on count of fishing boats landed, while there was a further stage for recording data on catch and other ancillary information. Here, also the landing centres formed a stratum in space but months were the strata in time. A month was considered more convenient as a time stratum, as it is sometimes necessary to study monthly trends and partly because of administrative convenience. The centres, days and time intervals were the successive stage of sampling units in observing boats. A fishing boat landed corresponds to the further stage of sampling in the case of observing catch. At each of the centres, four days were randomly selected in a month. Two days were allocated to observe count and the remaining two to observe catch.

Another extension survey was conducted during 1954-55 in order to finalise the technique for routine data collection. The coverage was the same as in the earlier survey. Here, the stratification used over space, was by fish-curing yard which was proved to be useful as it was convenient for field work. The month continued to be the stratum in time. Initially one single landing centre was selected at random. Four days were randomly chosen in a month and within a day the selected centre was kept under observation for three evenly spaced two hour intervals. In each interval one hour was meant for observing count of fishing boats landed and the other for recording the catch of a few selected boats. In 1955-56, some modification in the selection of centres was introduced. For a stratum, a fresh selection of centres was made for each day of the four days. In other aspects the design remained the same as that of the previous year.

On the basis of the experience gained in pilot surveys, a full-fledged survey was launched in an area of 200 miles of Travancore-Cochin coast consisting of 123 landing centres. Sukhatme et al. (1958) describe this sampling methodology and discuss the lack of a sampling frame in the absence of compulsory registration of boats. A group of contiguous landing centres is taken as a stratum in space and week was taken a stratum in time. The primary unit of sampling within each stratum was a landing centre. The centres were selected afresh every week. Each centre was kept under observation for two days (secondary stage units) selected at random out of the working days of the week. A day was divided into two clusters of three evenly spaced two hour intervals. The first comprises of intervals 0600-0800 hrs., 1000-1200 hrs., 1400-1600 hrs. and the second of intervals 0800-1000 hrs., 1200-1400 hrs. and 1600-1800 hrs. On each selected day, the field work was conducted for one cluster selected at random. In each two-hourly interval of a cluster, one hour was assigned for counting the number of boats landed and the other hour for recording the catch details.

The estimation procedure followed in the pilot surveys is given below.

The method of estimating total catch in the surveys conducted on the Malabar coast involved estimation of two factors, average catch per fishing boat $\bar{y}$, and the total number of boats operated M. These factors were used to arrive at the estimate of catch for any period or region, assuming that $\bar{y}$ and M are not correlated.

The total catch for a stratum and month was estimated by the product of (Total count for the stratum in a month) * (Average catch per fishing boat) = $\hat{M}\ \hat{\bar{y}}$.

Variance of total catch is determined using the formula

$$V(\hat{M}\ \hat{\bar{y}}) = \hat{\bar{y}}^2 V(\hat{M}) + \hat{M}^2 V(\hat{\bar{y}})$$

On the basis of the results obtained in the pilot surveys, the following broad conclusions were drawn about the different stages of design of large-scale sample survey for estimating fish production.

(i) Stratification: space stratification is to be followed. It is to be examined if grouping of centres according to amount of catch in the centres will improve the design. Time stratification is also to be introduced.

(ii) Size of the primary unit: In the two-fold stratification over space and time the primary unit may be (a) a centre-day, (b) a centre-group of days, (c) cluster of centre days, and (d) cluster of centres-group days. From the organisational point of view, while a field staff is put in charge of a stratum over space, the cluster of centres-day may not be possible. Among the rest a centre-day or a centre-group of days may be used as primary unit depending on field conditions.

(iii) Size of the ultimate unit: It is seen that the ultimate unit would be an interval of time (one hour, two hours etc.) in a day. The length of the interval has to be divided on the basis of statistical and field considerations. The sampling of interval within a day may however be done systematically.

(iv) Observation on count and catch: The pilot survey showed that observing count and catch in different time intervals entailed considerable loss of data, whenever there was no boat to be observed in the subsequent time interval. Therefore count and catch are observed simultaneously in the same interval.

Similar pilot surveys conducted in Madras coast, Andhra coast, South Canara and North Mumbai with minor modifications.

Census

The first Marine Fisheries Census was carried out in 1957-58. The initiation of data collection using a stratified multi-stage sampling scheme in the west coast of India was done in the year 1959. The second Marine Fisheries Census was carried out in 1961-62. During 1960-69, the survey scheme, has crossed the level of experimentation, entered into the phase of evaluation. During 1970-79, there was a spurt in the implementation of mechanization in the fisheries sector which yielded in dramatic increase in the quantum of data collected. The previous list of more spatially spread zones demanded a relook. As a few landing centres recorded very heavy landings due to increased harbour facilities and marketing avenues, they demanded for better representation in the scheme. Hence the move to treat such centres as exclusive zones known as, Single Centre Zones, was initiated and is being followed till date. The secondary stage units viz. fishing boats with different gears were segregated to have separate recordings so that the estimation can be done for each landing centre day for a given gear. A marine fisheries census was conducted in 1973, which augmented the information on sampling frame. Another census was conducted in 1980 in all maritime states except Maharashtra and the information was used to update the sampling frame. In early eighties, underlining the need of an organised conglomeration of data on marine living resources, the Planning Commission had suggested strengthening of Data Centre of CMFRI. A workshop on data acquisition was conducted in 1982, which delved on all the issues flagged out by the Planning Commission and based on the deliberations, a new scheme of acquiring, processing, analysing and storing of data and dissemination of information was charted out. The eighties witnessed significant impact of mechanisation. Motorisation of country crafts came in to existence. Hence the methodology demanded a thorough revision of gears region-wise and treated them as the domains. The selection criteria of crafts were modified. Eighties also ushered in high profile electronic computing, first through the mainframe concept and later as personal computing. This triggered development of more accurate computation methods which could be easily programmed for analysis and stored for future reference as electronic databases. After 1990, the methodology did not undergo major changes barring the deletion and addition to the existing list of landing centres. An All India Marine Fisheries Census was conducted in 2005 which facilitated updating of the sampling frame. It was a centrally sponsored scheme on Strengthening of Database and Information Networking for the Fisheries Sector under the aegis of Department of Animal Husbandry, Dairying and Fisheries (DAHD&F), Ministry of Agriculture, Government of India. Its reports on all India basis were presented to DAHD&F in July 2006. This census not only covered the details on the fish landing centres, crafts and gears but also the data on the marine fishermen population, their occupation, family status and the infrastructure available in the villages.

1.4 Marine Fisheries of Kerala

Kerala is a small state situated in the southwest corner of the Indian Peninsula between 8° 18' and 12° 48' north and 74° 52' and 77° 22' east. It is a narrow strip of lush green land bounded on the east by Western Ghats interspersed with rivers and on the west by the Arabian Sea. Kerala has got a long and unbroken coastline of about 590 kilometres, and is only 100 kilometres across at the widest point. The area of continental shelf of this coast is about 40,000 square km and the overlying waters are considered to be one among the most productive in the Indian waters. 9 out of the 14 districts in the Kerala state have Arabian Sea as their western border.

Fishing has been the traditional occupation for generations among people living in Kerala. Among the nine maritime states in India, Kerala occupies the foremost position in marine fish production. The contribution of Kerala fisheries to the economy of the country is substantial particularly with reference to food consumption, nutrition, employment and export. Although the Kerala coastline is only about one tenth of the coastline of India, it contributes more than 30% to the country's total marine fish production. As per the latest report of the Marine Products Export Development Authority, the marine products export from Kerala during 2005-2006 was 97,311 metric tonnes valued at Rs. 1257.65 crores constituting 17% in terms of value to Indian marine products export.

The marine fisheries of Kerala have progressed tremendously during the last five decades contributing significantly to the socio-economic welfare of the coastal rural folk and to the economy of the state. Marine fishing using artisanal tackles like boat-seines, shore-seines and Chinese dip nets are an age old tradition of the state. There have been qualitative and quantitative improvements in the scale and magnitude of fishing operations aided by scientific explorations and technological innovations as well as increasing demand for marine fish products both in domestic and international markets.

The progress of marine fisheries in the state has been quite eventful with each epoch witnessing different innovations of harvesting practices in the gears and craft. The mechanisation was experimented in the late fifties under the Indo-Norwegian Project (INP) by introducing trawlers. The INP project was undertaken under a joint agreement among the United Nations, the Government of Norway and the Government of India. (Kurien, 2000). When the INP started in 1953 there were around 38,000 active marine fishermen (Kurien 1985) and as per Marine fishery census 2005, there were 1,40,222 active fishermen in the year 2005.

The early sixties witnessed an important technological development in gear, the shift from cotton to nylon nets. For about three decades from the formation of the state of Kerala in 1956, fisheries development was associated almost totally with the catching and exporting of shrimp. The mid-sixties ushered in increased use of trawl fishing by mechanised craft targeted towards exploiting prawns, the major foreign exchange earner. Commercial purse-seining aimed at harvesting small pelagics such as oil sardine and mackerel was started during the late seventies. About two thirds of the marine fish landings of the state were accounted by the artisanal sector till 1979. In 1980 there were as many as 22 major craft-gear combinations used by the artisanal fishermen to harvest the resources of the coastal waters (Kurien and Willmann, 1982).

One of the most significant developments in the marine fisheries of the state has been the motorisation of country craft, which was initiated in the early eighties and gained momentum in the later half of eighties. There were significant changes in the gear used by the artisanal sector. Boat seine has been converted into the mini purse-seine (ring seine) and the country craft converted into the mini trawls. The introduction of ring seine net has transformed the marine fishery scenario of the state. The impact of ring seine and mini trawlers used in the artisanal fisheries in Kerala was examined by D'cruz (1998) and reported as harmful to the fishery. The mid-nineties witnessed the phenomenon of voyage and deep-water fishing by trawlers and gill-netters.

As per the South Indian Federation for Fishermen Society (SIFFS), (1992), the Kerala coastline was distinguished by at least 14 types of fishing crafts and at least 23 types of fishing gears. Although the technological innovations introduced from time to time have helped in augmenting the total production, they have also given rise to inter sectoral conflicts among various stakeholders. Serious concern was expressed during the mid-eighties about the sustainability of the exploited resources and ecosystem degradation allegedly due to increased fishing pressure by the mechanised sector. The artisanal sector whose sustenance depended upon the small pelagic resources and other near shore resources felt threatened by the reported incursions of the mechanised sector into their region of exploitation. This prompted the Government of Kerala instituting various committees over successive years to assess the status of the fishery and this has culminated into promulgation of Marine Fishery Regulation Act aimed at regulating and curbing fishing activity by certain gears and craft at certain clearly demarcated fishing zones. To protect the spawners from being over exploited and to safeguard the interests of the traditional fishermen, a partial ban on trawling was introduced in 1988 (Vijayan, et al. 2000). Thereafter, the ban on fishing by trawlers during the monsoon period was enforced. From the year 1994 onwards, the period of 'Trawl-Ban' during monsoon period was fixed for 45 days from 15th June to 29th July. The Status of marine fisheries in Kerala with reference to ban of monsoon trawling was described by Ammini (1999). There has been massive increase in ring seine operations after the implementation of the trawl ban. In addition to this, the fishing pattern also underwent changes through extension of fishing grounds to relatively deeper zones and stay over or voyage fishing aided by state-of-the-art electronic equipments for communication, position fixing and resource detection.

In Kerala, the Marine Fisheries Census 2005 was carried out in all the nine coastal districts during April-June, 2005. There were around 2,24,606 people depending on fisheries for their livelihood. Of the 222 fishing villages in Kerala, the largest number is in Trivandrum district, 42 and the least is in Kannur district, 11. There are 178 landing centres in the state. Thiruvananthapruam district has the maximum number of landing centres, 50 and Kannur has the minimum, 11.

The total marine fishermen population in the state is about 6,02,234 of which 23% are engaged in actual fishing. Among those involved in actual fishing, 88.5% were engaged in full time fishing. 7.5% part time and 4% occasional. Full time fishermen were higher in Thiruvananthapuram district.

Trawlers (72%), ring-seiners (8%) and gill-netters (7.8%) were the main crafts of the mechanised sector. There were 29,177 crafts in the fishery of which 5,504 were mechanised; 14,151 were motorized and the rest non motorised. Kerala marine fisher folk owned 19,173 crafts out of which 7% were mechanised 44% were motorized and the remaining 49% were non-motorized crafts.

Important gears of Kerala were gillnets, hooks and lines, troll lines, drift nets, seines and trawl nets. Sharing pattern is more visible in seines, trawl nets and drift nets. Nearly 66% of the fisher folk families involved in fishing possessed neither craft nor gear. There were 414 curing yards, 320 ice factories, 153 peeling sheds, 112 boat yards and 56 freezing plants in the fishing villages of Kerala.

1.5 Objectives of the Study

The sampling methodology currently adopted by the CMFRI to estimate the marine fish landings in Kerala is based on stratified two stage sampling scheme, the stratification is done over space and time. The Kerala coastline is divided into several geographic contiguous zones. Each zone is taken as a space stratum and they are made by combining the adjacent landing centres. The stratification over time is by calendar month. A combination of landing centre and day called as landing centre day forms the primary stage unit and the fishing boats land on a landing centre day forms the secondary stage units.

In early fifties, a three stage sampling scheme was followed for estimation of marine fish landings in which a landing centre, a time interval of 20 minutes and fishing boats were the primary, secondary and ultimate sampling units respectively. Later on there have been changes and improvements in the sampling scheme from time to time in view of practical contingencies. One of the main changes has been in the selection process of space-time units i.e., a combination of landing centre and a day forming the primary stage unit. 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 by Kutty et.al (1973) there had been no attempt to evaluate the sampling design of CMFRI in terms of the precision of the estimates and deriving optimum sample size. In this study an attempt is made to review the existing sampling design in tune with the rapid changes in the fishery sector. The present investigation is proposed with the following objectives:

1. To evaluate the existing sampling design followed by CMFRI for estimation of marine fishery resources in Kerala.
2. To suggest improvement in the sampling design / estimate.
3. To estimate the optimum sample size to evaluate the catch and effort data.

1.6 Review of Literature

In this section, a brief review of the literature related to fishery surveys, connected sampling designs, methods of estimation are provided.

Fishery Surveys

The earliest reference to estimates of marine fish catch in India is seen in the Report on Marketing of fish in the Indian Union (Government of India, 1951) which also reports that the data were not based on any scientifically planned surveys but mostly on trade enquiries and similar other evidences (CSO, 1961). Bal and Banerji (1951) gave an account of the efforts made by the CMFRI in developing such a survey. Between 1950 and 1956 the ICAR initiated a number of pilot surveys of various designs in different regions of the country and the details are given in ICAR technical bulletin (ICAR, 1965). The estimation of yield from exploited marine stocks with reference to South East Asia was described by Banerji and Chakraborty (1972).

CMFRI conducted a workshop in 1982 to review the system of collection, collation and analysis and dissemination of data on marine living resources in the country (CMFRI, 1983). The sampling design that was followed up to 1970's was explained by Kutty et al. (1973) and evaluation of the design was done to see whether any improvement in the sampling procedure is possible by increasing the number of survey staff. The mode of collection during the late 1970's and early part of 1980's were described by Jacob et al. (1983). Later, the mode of collection underwent slight change with respect to selection of crafts and the modified scheme was given by Alagaraja (1984). Srinath et al. (2005) described in detail the existing sampling methodology followed by CMFRI for estimating marine fish landings and the expended fishing effort. The progress of the development of sampling scheme was given by Srinath and Jayasankar (2007).

Gulland (1955) discussed an analysis of samples from commercial landings of the English trawl fishery. Panse and Sastry (1960) carried out sample surveys in United Arab Republic for the improvement of fisheries statistics in that country. Tomlinson (1971) described methods for sampling commercial fisheries for small pelagic fish using two-stage sub-sampling with primary units of unequal size sampled with equal or unequal probabilities. Brander (1975) gave the guidelines for collection and compilation of fishery statistics. Statistical procedure to analyse data from the Pacific Halibut Fishery of the United States and Canada are discussed by Southward (1976) who presents a double-sampling procedure based on time and area stratification for estimating the age and length distribution in landings. For estimating the catch at age of rockfish on the west coast of California, Sen (1986) developed a two-stage sampling plan with boat trips as first stage units post-stratified into categories and clusters sub-sampled from each category. Sen (1990) also developed a cost-effective sampling plan for obtaining reliable estimates of annual catch by recreational fishermen in Hawaii and the effort expended by fishing method for some of the important management species. Papaconstantinou et al. (2002) presents the design of an integrated sample survey system in Greece for the collection of multiple fisheries data required for fisheries management. Stamatopoulos (2002) discusses the methodological and operational concepts in fishery data collection systems. Miller et al. (2007) presented an approach for determining sampling fractions and sample sizes for each stratum within a stratified sampling design that is optimal with respect to multiple parameters that may be heterogeneous in nature.

Sampling Designs

Several text books such as Hansen et al.(1953), Desraj (1971), Murthy (1967), Cochran(1977), Sukhatme et al.(1997), Krishnaiah and Rao (Eds.) (1988) Särndal et al. (1992) and Thompson (2000) describe the wide variety of efficient sampling designs together with the appropriate estimation methods.

The general theory for sub-sampling from finite populations has been developed through the important contributions of Hansen and Hurwitz

(1943, 1949), Mahalanobis (1946, 1952), Sukhatme (1947, 1950), Yates (1949), Sukhatme and Panse (1951), Sukhatme and Narain (1952), Singh (1954), Durbin (1967), Sukhatme and Sukhatme (1976) and others. Procedures for estimation of various population parameters were developed for multistage design in conjunction with stratification and other sampling schemes such as selection with equal or unequal probabilities, selection with or without replacement etc.

Cochran (1946) considered a model for auto-correlated populations with a view to compare systematic sampling with stratified sampling and simple random sampling procedures. With the importance of this model in view, an attempt to deal with the problem of sampling from two dimensional populations was made by Quenouille (1949). Das (1950) gave an independent approach to the problem of two dimensional population in the context of systematic and stratified random sampling. Sukhatme et al. (1958) discussed the technique of two dimensional population for the estimation of the catch of marine sea fish in India.

Singh and Gupta (1980) estimated the production of vegetable crops where harvesting is spread over a period of about 2-3 months. They suggested that it is reasonable to assume that the number of observations made on a unit is a discrete random variable and its distribution may be assumed to be characterised by a Truncated Poisson distribution. On the basis of this, they obtained unbiased estimate of the population mean, the variance of the estimator and the estimate of variance of the estimated mean.

Kumar (1981) suggested a sampling plan for two dimensional studies using random sampling for selection of fields and systematic sampling over time. Mahajan (1984) investigated the use of successive sampling in two dimensional populations and developed suitable estimation theory for sampling on successive occasions.

Estimation of Variance

Anderson and Bancroft (1952), Gower (1962) and Gates and Shiue (1962) and Mahamunulu (1963) gave the expected mean squares for unequal sampling from infinite nested population. Bennett and Franklin (1954) considered finite nested populations, but only for balanced sampling. Gaylor and Hartwell (1969) gave a single unified procedure for obtaining the expected mean squares in nested populations, where a balanced or unbalanced random sample is taken from a finite or infinite number of levels for each classification. Khuri (2000) provides a comprehensive coverage of the literature on designs for estimating variance components, and a review of recent applications of such designs in genetics, statistical process control, and quality improvement. In addition, recent methods of estimation of variance components and model forms, other than the linear, are discussed.

Post-Stratification

Hansen et al. (1953) was the first to discuss the concept of post stratification. Williams (1962) 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 Sukhatme and Sukhatme (1976), Murthy (1967) and Cochran (1977) and others.

Fuller (1966) developed small sample estimator for two post-strata and compared with pooling or collapsing procedures commonly employed in practice. Holt and Smith (1979) showed that neither the post-stratified estimator nor the sample mean is uniformly best in all situations but empirical investigations indicate that post-stratification offers protection against unfavourable sample configurations and should be viewed as a robust technique.

A post-stratified cluster sample design was proposed by Akar and Sedransk (1979). They suggested an estimator $\hat{R}$ of the finite population ratio. Post-stratification in unistage cluster sampling on the basis of the elements of selected clusters has been discussed by Mehrotra et al. (1984). They have demonstrated empirically that the suggested procedure not only provides estimates of the character under study according to the strata variable but also improves the precision of the overall estimate compared to the usual cluster sampling procedure. Sethi and Srivastava (1987) have developed ratio estimators with post-stratified design in order to overcome the constraints imposed by the assumption viz., sample mean of the auxiliary character should always be less than twice the population mean of the character on the application of ordinary ratio estimation theory. Pfeffermann and Krieger (1991) have proposed a new regression type estimator which accounts for different regression relationships in various strata but no longer depends on the unknown strata means and sizes under post-stratification.

Mehrotra (1993) has given a scheme for post-stratification in two-stage sampling on the basis of the sample second stage units. It has been empirically demonstrated that the scheme not only provides estimates of the study character as per the strata variable but also improves the precision of the estimate pooled over the strata compared to the conventional non-stratified two-stage procedure. Kumar (1989) has extended the above schemes of post-stratification in uni-stage unequal cluster sampling and post-stratification in two stage sampling on the basis of sample second stage units to two and three stages respectively. He has also discussed post-stratification for two stage sampling with probability proportional to size. Narang (1994) has given schemes for estimation of population total in a two-stage post-stratified design.

1.7 Plan of the Thesis

The thesis consists of SIX chapters. Each chapter begins with a brief introduction highlighting the chapter contents. In the earlier sections of this chapter a brief description about the CMFRI and its efforts to collect marine fishery data, the background and objectives of this study, a brief review of the related literature are given. In the second chapter, a critical evaluation of the sampling design currently followed by CMFRI is done. Section 2.2

describes the sampling design and the procedures of estimation of total landings and its variance. Based on the CMFRI data for the years 2004-2006, the sampling fraction and the estimates of landings in each zone are described in section 2.3. Critical evaluation and the main limitations of the existing design are described in section 2.4. A few remedial measures are indicated in section 2.5.

The fluctuations and trends in the fish landings data over the years are discussed in chapter 3. Section 3.2 describes how the fisher folk population and their fishing activity varied over time. In section 3.3 the trends in the assemblage wise and sector wise landings are described by fitting suitable trends to the data. In section 3.4 the trends in the landings, boats operated and the landings per boat are discussed in detail by fitting trends for monthly data based on 2 term, 3 term, 4 term and 6 term moving averages. Comparisons of the period wise landings at the five single centre zones are made in section 3.5. Section 3.6 concludes the findings in the analysis made in the chapter.

In chapter 4, estimation of the components of variance due to each stage of sampling in the fish landing data using the nested model technique is discussed. The general linear model for ANOVA of an unbalanced nested design and its analysis are described in section 4.2. Describing the nested structure of the marine fishery data, analysis of the data using a three stage model is described in section 4.3. Zone wise estimates of variance components for the period 2004-06 are described in section 4.4. The chapter ends with a brief conclusion of the findings.

In chapter 5 we introduce two modified sampling designs based on post stratification of the data – one applicable to single centre zones and the other for multi-centre zones. The first new design developed for single centre zones is the same as the two stage sampling currently followed by CMFRI with the only modification that the data is post-stratified according to the observed gear types is described in section 5.2. Illustration of this design is also discussed. The second new design developed is a three stage design which retains the two dimensional structure of the population over space and time and adopts post-stratification based on gear types at the third stage and is described in section 5.3. The chapter concludes by highlighting the effectiveness of the proposed designs.

A more scientific, structurally and operationally simple design is developed in chapter 6. The new design – a two stage probability proportional to size sampling design applicable to multi-centre zones is developed in section 6.2. The optimum sample sizes for estimation of the fish landing data under different sampling designs are described in section 6.3.

A final conclusion together with some recommendations on the basis of the findings of the study is given at the end. The list of references is also included.

CHAPTER II

CMFRI Sampling Design in Kerala – An Evaluation

2.1 Introduction

The data collection scheme of CMFRI, based on a stratified two stage sampling, involving space-time stratification, was first put into operation in Kerala state in the middle of 1959. The vast experience gained in the collection of marine fish catch statistics and the results of the several pilot surveys conducted, described in sections 1.3 and 1.4, have a significant role in the development of the sampling design currently followed by the Institute. In this chapter a critical evaluation of the sampling design currently followed by CMFRI is done. Section 2.2 describes the currently used sampling design and the procedures of estimation of total landings and its variance. Tables indicating the sampling fraction achieved in each zone are given in section 2.3. Critical evaluation and the main limitations are described in section 2.4. A few remedial measures are indicated in the last section.

2.2 CMFRI Sampling Design

The currently employed sampling design to estimate the annual marine fish landings in Kerala is described below.

The sampling design currently adopted by the CMFRI to estimate species-wise/gear-wise landings is based on stratified two-stage sampling technique, the stratification is done over space and time. By combining the adjacent landing centres out of the 177 landing centres in the state, 172 are divided into 10 geographically contiguous zones containing 11-26 landing centres. The remaining 5 landing centres with relatively high intensity of fishing activity are regarded as single centre zones. The multi-centre zones are further divided into strata based on the intensity of fishing activity. The district wise distribution of landing centres along the coastline of Kerala and their zonal distribution are given in tables 2.1(a) and 2.1(b) respectively. The geographical distribution of the zones is given in fig 2.1.

Table 2.1(a) Distribution of the landing centres over the coastline districts

District	Length of Coastline (in km)	Number of landing centres
Thiruvananthapuram	78	50
Kollam	37	18
Alappuzha	82	13
Ernakulam	46	12
Thrissur	54	19
Malappuram	70	12
Kozhikode	71	25
Kannur	82	11
Kasaragod	70	17
Total	590	177

Figure 2.1 Geographical distribution of the fishing zones in Kerala

Table 2.1(b) Zonal distribution of landing centres

Serial No.	Name of zone	Zone label	Number of landing centres	Number of strata
1	Thiruvananthapuram1	K1	25	3
2	Thiruvananthapuram2	K2	25	3
3	Kollam	K3	16	3
4	Alappuzha	K4	13	4
5	Ernakulam	K5	9	1
6	Thrissur	K6	19	3
7	Malappuram	K7	12	3
8	Kozhikode	K8	25	3
9	Kannur	K9	11	3
10	Kasaragod	K10	17	3
11	Neendakara	K11	1	1
12	Sakthikulangara	K12	1	1
13	Cochin	K13	1	1
14	Munambam	K14	1	1
15	Vypin	K15	1	1

Each zone is regarded as a stratum in space. The stratification over time is by calendar month. A zone and a calendar month constitute a space-time stratum. If in a zone, there are 25 landing centres and there are 30 fishing days in the month; we get 25 x 30 = 750 landing centre days which constitute the primary stage units (PSU). The fishing boats that land on a landing centre day forms the second stage units (SSU).

The introduction of space–time stratification in the sampling methodology becomes necessary as the fish population is supposed to vary with respect to both space and time. The stratification is intended to reduce the variance in the sample estimates. The fish landings are found to vary considerably among the landing centres in a multi-centre zone, especially in different seasons and hence a zone is further stratified as major, minor and very minor centres etc. The centres in which either mechanised boats or 100 or more non-mechanised/motorised boats are operating are considered as major centres. Similarly other strata are defined based on the number and type of fishing boats operating.

2.2.1 Selection of Primary Stage Units

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 cluster 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. From among the total number of landing centers in a zone, 9 centres are selected with replacement and allotted to the 9 cluster days selected as described earlier. These 9 days are evenly distributed among the strata in case of multi-centre zones. A landing centre day which is the PSU is the 24 hour duration from noon of the first day to the noon of the following day.

2.2.2 Period of Observation

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 landing centre day. The enumerator will be present through out the periods 1 and 2 at the centres. The data on landings during period 3 (night landings) is usually collected from the landing centre by enquiry on the following day morning. The sum of the observations on the 3 periods contribute the data for the landing centre day.

2.2.3 Selection of Second Stage Units and Recording of Fish Landings

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 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 25 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 procedures of selection of the landing centre days and the boats landed on the selected day for single centre zones are the same as in the case of a stratum in a multi-centre zone.

2.2.4 Administration of the Survey

Plan of operation

The survey staff is given 10-12 weeks training course immediately after recruitment and is posted to the survey centres. They are permanent employees. Each survey centre is housed in 1-2 room accommodation and each centre is provided with literature connected with the identification of fish, a reference collection of local fish species, crustaceans and molluscs, field notebooks and registers. The programme of work for the following month is carefully designed by the staff of Fishery Resources Assessment Division at the CMFRI headquarters. Generally one field staff is allotted to each zone to collect the fish landings data. At the end of every month, the survey staff receives the programme of work for the next month by post, that includes the names of landing centres to be observed and details such as dates and time for observations at each landing centre. The field staff are instructed to send the data collected during every month to reach the Institute's headquarters at least by the end of first week of the subsequent month.

Supervision of data collection

Surprise inspections are carried out by the supervisory staff of the Institute and the enumerators are inspected while at work in the field and their field notebooks and diaries are scrutinised. The estimated zonal landings are always compared with the previous year's survey figures, and if any variation which cannot be explained is observed, the technique of interpenetrating sub-samples is adopted to detect observational errors. Observational errors are rarely encountered and when confirmed, the field staff is either called back to the headquarters for giving intensive training or he is replaced. Zonal workshops are held periodically to review the progress of work and update the sampling frame and to impart refresher courses to the field staff.

Errors due to non-response, their magnitude and control

Non-response occurs when the regular field staff is not available to observe the centre-day included in the sample. Usually, arrangements are made at the Headquarters/Research/Regional Centre to minimise the non-response.

2.2.5 Analysis of Data

In the existing sampling methodology, the interest is to estimate gear-wise, species-wise landings for the state in a month, fishing effort according to different types of fishing boats and also in terms of man hours. The analysis is carried out at CMFRI headquarters. Before the data is processed for analysis it will be ensured that the data collection is made as per the approved schedule, by checking the appropriate proformae. The responsibilities and functions of staff at the headquarters are data coding, estimation and database management. The data analysis is computerised and estimates are made using the software developed by the Fishery Resources Assessment Division of the Institute. The processed data are again counter- checked for errors. When discrepancies are detected, the estimation procedure is scrutinised in detail.

Monthly estimate for a zone

Since there are two types of zones, single centre and multi-centre zones, the method of estimation are described in respect of each of them separately.

(a) Single Centre Zone

In the case of single centre zones, there is no stratification over space.

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. ($k = 1,2,...,T_i; i = 1,2,...,N$; $p = 1,2,3$). 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. ($l = 1,2,...,m_{ikp}$)

The raw data collected by the existing sampling design can be represented in a tabular form given in Table 2.2.

Table 2.2 Fish landings data as a two-way table with n days and T_i gears during p^{th} period of observation

LCD	Gear				
	1	…	k	…	T_i
1	$y_{11p1}, y_{11p2}, \ldots, y_{11pm_{11p}}$	…	$y_{1kp1}, y_{1kp2}, \ldots, y_{1jkm_{1kp}}$	…	
2	$y_{21p1}, y_{21p2}, \ldots, y_{21pm_{2p1}}$		$y_{2kp1}, y_{2kp2}, \ldots, y_{2kpm_{2kp}}$	…	
⋮	⋮		⋮	⋮	
i	$y_{i1p1}, y_{i1p2}, \ldots, y_{i1pm_{i1p}}$	.	$y_{ikp1}, y_{ikp2}, \ldots, y_{ikpm_{ikp}}$	…	
⋮	⋮	⋮	⋮	:	
n	$y_{n1p1}, y_{n1p2}, \ldots, y_{n1pm_{n1p}}$	…	$y_{nkp1}, y_{nkp2}, \ldots, y_{nkpm_{nkp}}$	…	

LCD- landing centre day

Let $\bar{y}_{ikp}$ 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

$$\bar{y}_{ikp} = \frac{1}{m_{ikp}} \sum_{l=1}^{m_{ikp}} y_{ikpl} \qquad \text{-----(2.1)}$$

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

$$\hat{Y}_{ikp} = M_{ikp}\, \bar{y}_{ikp} \qquad \text{-----(2.2)}$$

The estimates of landings by different types of gear on the selected days are given in Table 2.3. It can be seen that, each cell of the Table 2.2 was modified by replacing the observed landings with the estimates of landings for each type of gear during the p^{th} period of observation in Table 2.3.

Table 2.3 Adjusted data during the p^{th} period of observation

Day	Gear					Sum
	1	...	k	...	T_i	
1	$Y_{11p} = M_{11}\bar{y}_{11p}$	...	$Y_{1kp} = M_{11}\bar{y}_{1kp}$	...		$Y_{1.p}$
2	$Y_{21p} = M_{21}\bar{y}_{21p}$	...	$Y_{2kp} = M_{2k}\bar{y}_{2kp}$	...		$Y_{2.p}$
⋮	⋮	⋮	⋮	⋮		⋮
I	$Y_{i1p} = M_{i1}\bar{y}_{i1p}$	:	$Y_{ikp} = M_{ik}\bar{y}_{ikp}$	:		$Y_{i.p}$
⋮	⋮	⋮	⋮	⋮		⋮
N	$Y_{n1p} = M_{n1}\bar{y}_{n1p}$	:	$Y_{nkp} = M_{nk}\bar{y}_{nkp}$	:		$Y_{n.p}$
Sum	$Y_{.1p}$	...	$Y_{.kp}$	...		$Y_{..p}$

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

then

$$\hat{Y}_{ik.} = \sum_{p=1}^{3} \hat{Y}_{ikp} \qquad \text{-----(2.3)}$$

The night landings (p=3) are obtained by enquiry and usually estimated from the total number of each type of boat landed and average catch per boat.

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

$$\hat{Y}_{i..} = \sum_{k=1}^{T_i} \hat{Y}_{ik.} \qquad \text{-----(2.4)}$$

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

$$\hat{Y}_{.k.} = N\hat{\bar{Y}}_{.k.}, \qquad \text{-----(2.5)}$$

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

The estimated total landings for the month over all distinct gear types T is given by,

$$\hat{Y} = \sum_{k=1}^{T} \hat{Y}_{.k.} \qquad \text{-----(2.6)}$$

Estimate of variance of total landings

Assuming that the variance between gears and between boats of the same gear are negligible (Sukhatme *et al.* 1958) within a selected landing centre day, the variance of the total landings is given by

$$\hat{V}(\hat{Y}) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)\frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{Y}_{i..} - \bar{Y}_{...}\right)^2 \qquad \text{-----(2.7)}$$

where $\hat{Y}_{i..}$ is total landings given by equation (2.4) and $\bar{Y}_{...}$ is the average landings per landing centre day.

(b) Multi-centre zone with stratification

Let the zone with N centres be divided into L strata (in practice a zone is divided into 3 or 4 strata depending on the intensity of fishing operations). The number of fishing days in a month D is regarded as the same for all strata.

Let N_h be the number of centres in the h^{th} stratum of a given zone and n_h be the number of landing centres selected belonging to h^{th} stratum $(h = 1,2,...L)$. Let M_{hikp} be the total number of boats of gear type k landed during p^{th} period of observation on the i^{th} selected landing centre day in the h^{th} stratum. $(k = 1,2,...,T_i;\ i = 1,2,...,N_h D;\ p = 1,2,3)$. Let m_{hikp} be the number of selected boats of gear type k during the p^{th} period of observation on the i^{th} landing centre day in the h^{th} stratum, $(l = 1,2,...,m_{hikp})$.

Let y_{hikpl} be the quantity of fish landed by the l^{th} boat during p^{th} period of observation by the k^{th} gear type on i^{th} selected day in the h^{th} stratum.

Let $\bar{y}_{hikp}$ be average quantity of fish landed during p^{th} period of observation by the k^{th} type unit on i^{th} selected day in the h^{th} stratum and it is given by

$$\bar{y}_{hikp} = \frac{1}{m_{hikp}} \sum_{l=1}^{m_{hikp}} y_{hikpl} \qquad \text{-----(2.8)}$$

Let $\hat{Y}_{hikp}$ be the estimated total landings during p^{th} period of observation by k^{th} type unit on i^{th} selected day in the h^{th} stratum and it is given by

$$\hat{Y}_{hikp} = M_{hikp}\, \bar{y}_{hikp} \qquad \text{-----(2.9)}$$

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

$$\hat{Y}_{hik.} = \sum_{p=1}^{3} \hat{Y}_{hikp} \qquad \text{-----(2.10)}$$

The estimated total landings ($\hat{Y}_{hi.}$) for the i^{th} landing centre day in the h^{th} stratum is obtained as

$$\hat{Y}_{hi..} = \sum_{k=1}^{T} \hat{Y}_{hik.} \qquad \text{-----(2.11)}$$

The estimated total landings by the k^{th} gear type in the h^{th} stratum of a zone ($\hat{Y}_{h.k.}$) is given by

$$\hat{Y}_{h.k.} = DN_h\, \hat{\bar{Y}}_{h.k.}, \qquad \text{-----(2.12)}$$

where $\hat{\bar{Y}}_{h.k.} = \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{Y}_{hik.}$ is estimated average landing by the k^{th} gear type per landing centre day in the h^{th} stratum.

The estimated total landings in the multi-centre zone for the month ($\hat{Y}'$) is given by

$$\hat{Y}' = \sum_{h=1}^{L}\sum_{k=1}^{T}\hat{Y}_{h.k.} \qquad \text{-----(2.13)}$$

Estimate of variance of total landings

The variance is given by

$$\hat{V}(\hat{Y}') = \sum_{h=1}^{L} N_h^2 D^2 \left(\frac{1}{n_h} - \frac{1}{N_h D}\right) \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{Y}_{hi..} - \bar{Y}_{h...}\right)^2 \qquad \text{-----(2.14)}$$

where $\hat{Y}_{hi..}$ is total landings given by equation (2.11) and $\bar{Y}_{h...}$ is the average landings per landing centre day in the h^{th} stratum. The standard error of the estimate can be found out from the above formula.

Monthly estimate for the state

By pooling the zonal estimates as given by (2.6) and (2.13), both monthly and yearly estimates for the state as a whole can be obtained. Let Z denote the number of zones in the state (including single centre zones). Denoting the zone totals as given by (2.6) and (2.13) by $\hat{Y}_z$, the total landings for the state as $\hat{Y}''$, we get

$$\hat{Y}'' = \sum_{z=1}^{Z}\hat{Y}_z \qquad \text{-----(2.15)}$$

Estimate of fishing effort

The fishing efforts are usually expressed by

(i) the number of unit operations by a craft-gear combination (unit),

(ii) the fishing hours expended by the boat during the month, and

(iii) the man-hours expended by the boats during the month.

The estimation procedures for (i) and (ii) are described below while the procedure (iii) is exactly the same as (ii).

(i) Estimation of the number of unit operations

Let M_{hikp} be the number of boats landed during the p^{th} period of observation of k^{th} gear type on the i^{th} landing centre day in the h^{th} stratum in a zone (in the case of single centre zones, h=1). Then, $M_{ik.}$ the total number of k^{th} type of units during i^{th} day of observation is

$$M_{ik.} = \sum_{h=1}^{L}\sum_{p=1}^{3} M_{hikp} \qquad \text{----(2.16)}$$

The estimated number of unit operations of k^{th} type of unit $\hat{U}_k$ for a month is given by

$$\hat{U}_k = \frac{ND}{n}\sum_{i=1}^{n} M_{ik.} \qquad \text{-----(2.17)}$$

(ii) Estimated effort in fishing hours

Let f_{hikpl} be the effort expended in actual fishing hours by the l^{th} selected boat during p^{th} period of observation of the k^{th} gear type observed on the i^{th} landing centre day in the h^{th} stratum (in the case of single centre zones, h=1).

Let $\hat{f}_{hikp}$ be the estimated effort expended by the k^{th} gear type during p^{th} period of observation on i^{th} selected landing centre day in the h^{th} stratum, then

$$\hat{f}_{hikp} = M_{hikp} \times \bar{f}_{hikp}, \qquad \text{-----(2.18)}$$

where $\bar{f}_{hikp}$ is the average effort expended by the k^{th} gear type during p^{th} period of observation on i^{th} selected day in the h^{th} stratum and it is given by

$$\bar{f}_{hikp} = \frac{1}{m_{hikp}} \sum_{l=1}^{m_{hikp}} f_{hikpl}$$

Let $\hat{f}_{hik.}$ be the estimated total effort expended by k^{th} gear type on i^{th} selected day in the h^{th} stratum, then

$$\hat{f}_{hik.} = \sum_{p=1}^{3} \hat{f}_{hikp} \qquad \text{-----(2.19)}$$

The estimated total effort $(\hat{f}_{hi..})$ for the i^{th} landing centre day in the h^{th} stratum is obtained as

$$\hat{f}_{hi..} = \sum_{k=1}^{T} \hat{f}_{hik.} \qquad \text{-----(2.20)}$$

The estimated total effort ($\hat{f}_{h.k.}$) by the k^{th} gear type in the h^{th} stratum for the month is obtained as

$$\hat{f}_{h.k.} = ND \times \hat{\bar{f}}_{h.k.} \qquad \text{-----(2.21)}$$

where $\hat{\bar{f}}_{h.k.} = \frac{1}{n}\sum_{i=1}^{n}\hat{f}_{hik.}$ is the average effort expended by the k^{th} gear type in the h^{th} stratum for the month.

The estimated total effort ($\hat{f}$) by all gear types in a month in a multi-centre centre zone is given by

$$\hat{f} = \sum_{h=1}^{L}\sum_{k=1}^{T}\hat{f}_{h.k.} \qquad \text{-----(2.22)}$$

The minimum sample size

An important question that had to be answered in planning a sample survey is about the minimum size of the sample required for estimating the population parameter with a specified precision. The precision is usually specified in terms of the margin of error permissible in the estimate and the coefficient of confidence with which one wants to make sure that the estimate is within the permissible margin of error. Thus, if the error permissible in the estimate of the mean is say, $\varepsilon\,\bar{y}_n$ and the degree of assurance desired is $1-\alpha$, then the size of the sample is determined so that

$$P\left\{\left|\bar{y}_n - \bar{Y}_N\right| \geq \varepsilon\,\bar{y}_n\right\} = \alpha \qquad \text{-----(2.23)}$$

where $\bar{Y}_N$ is the true mean and ε is a very small positive number. From the above condition the minimum required size 'n' of the sample is given by Sukhatme and Sukhatme (1976) is as follows.

$$n = \frac{\dfrac{t^2_{(\alpha,n-1)}}{\varepsilon^2} \dfrac{S^2}{\bar{y}_n^2}}{1 + \dfrac{1}{N} \dfrac{t^2_{(\alpha,n-1)}}{\varepsilon^2} \dfrac{S^2}{\bar{y}_n^2}}, \qquad \text{-----(2.24)}$$

where y is the characteristic under consideration,

N is the number of sampling units in the population,

$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$, is the sample mean,

S^2 is the population mean square and

$t_{(\alpha,n-1)}$ is the value of the student t- distribution corresponding to the level of significance α for $(n-1)$ degrees of freedom.

The above procedure is applied to a zone without stratification to find out the required sample size at various margin of error.

2.3 Zonal Sampling Fractions and Estimates of Landings

The zone wise and month wise sampling fractions attained in the CMFRI sample survey over the period 2004-06 are given in this section. The data were analysed for each stratum (zone-month) to determine the sampling coverage and the optimum sample size. Tables 2.4(1) to (15) give the sampling fractions achieved in the 15 zones during 2004-06. The minimum sample size required to estimate the zone totals within a margin of 10, 15 and 20% errors in the case of three single centre zones are given in table 2.5. The zone wise findings are described below the tables. The months for which data were not available are left blank in the tables.

Zone K1

Table 2.4(1) Zonal sampling fractions during 2004-06 – Zone K1

Month	2004			2005			2006		
	ND	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$
Jan	775	4	0.52	775	4	0.52	775	5	0.65
Feb	725	1	0.14	700	4	0.57	700	4	0.57
Mar	775	5	0.65	775	3	0.39	775	6	0.77
Apr*	750	4	0.53	-	-	0.69	750	4	0.53
May*	775	4	0.52	-	-	0.54	775	4	0.52
Jun	720	3	0.42	720	5	0.54	750	4	0.53
Jul	744	5	0.67	744	4	0.53	744	5	0.67
Aug	744	3	0.40	744	4	0.52	744	4	0.54
Sep	750	4	0.53	750	4	0.40	750	5	0.67
Oct	775	4	0.52	775	4	0.52	775	4	0.52
Nov	750	4	0.53	750	3	0.52	750	4	0.53
Dec	650	4	0.62	700	4	0.57	775	4	0.52

*The data for April-May 2005 were not available because the field staff were engaged in data collection for marine fishery census during this period.

The zone K1 consisting of 25 landing centres divided into three strata consisting of 3, 12 and 10 centres. The landings of the zone contribute nearly 4% of the total landings of Kerala during the period 2004-06. From the Table 2.4(1) it could be seen that the sampling fraction is far less than 1%. It is seen that the third stratum was very rarely observed. During the months of June, July and August, only stratum 1 was observed.

Zone K2

Table 2.4(2) Zonal sampling fractions during 2004-06 – Zone K2

Month	2004			2005			2006		
	ND	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$
Jan	775	4	0.52	775	4	0.52	775	4	0.52
Feb	725	3	0.41	700	3	0.43	700	3	0.43
Mar	775	4	0.52	775	3	0.39	775	4	0.52
Apr*	750	4	0.53	-	-	-	750	4	0.53
May*	775	5	0.65	-	-	-	775	3	0.39
Jun	750	2	0.27	750	3	0.40	510	5	0.98
Jul	527	4	0.76	775	3	0.39	527	3	0.57
Aug	527	3	0.57	527	4	0.76	527	5	0.95
Sep	750	4	0.53	750	2	0.27	750	4	0.53
Oct	775	4	0.52	775	5	0.65	775	4	0.52
Nov	750	4	0.53	750	2	0.27	750	4	0.53
Dec	650	3	0.46	775	5	0.65	775	4	0.52

The zone K2 consisted of 25 fish landing centres with three strata consisting of 8, 12 and 5 landing centres respectively. During the period 2004-06, nearly 3% of the total landings of Kerala were contributed by the zone. The sampling fraction achieved was less than 1. The third stratum was very rarely observed. In certain months, only stratum 1 was observed. The maximum landings occurred during July-September in all the years except in 2004. The very high landings in March 2004 was due to the use of large number of non-mechanised shore-seine units.

Zone K3

Table 2.4(3) Zonal sampling fractions during 2004-06 – Zone K3

Month	2004			2005			2006		
	ND	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$
Jan	465	4	0.86	744	4	0.54	744	8	1.08
Feb	464	4	0.86	672	4	0.60	672	7	1.04
Mar	496	4	0.81	744	5	0.67	744	7	0.94
Apr*	480	2	0.42	-	-	-	720	8	1.11
May*	465	3	0.65	-	-	-	682	5	0.73
Jun	420	6	1.43	720	9	1.25	660	13	1.97
Jul	496	11	2.22	744	17	2.28	682	16	2.35
Aug	496	4	0.81	744	2	0.27	744	10	1.34
Sep	480	6	1.25	720	6	0.83	720	7	0.97
Oct	496	5	1.01	744	5	0.67	744	8	1.08
Nov	480	5	1.04	720	6	0.83	720	5	0.69
Dec	416	4	0.96	744	6	0.81	744	7	0.94

Zone K3 covers the coastline of Kollam district. There were 16 centres grouped into three strata with 4, 4 and 8 centres respectively. The landings of this zone accounted for nearly 4% of the total landings of Kerala during the period 2004-06. The stratum 3 was observed only in the year 2004. In certain months, only stratum 1 was covered. It could be seen that the maximum sampling fraction were around 1% itself except during the months of June and July. The slightly higher sampling fraction in these months were due to deployment of more field staff there due to ban in the neighbouring zones. The landings were high during May-September all the years.

Zone K4

Table 2.4(4) Zonal sampling fractions during 2004-06 – Zone K4

Month	2004			2005			2006		
	ND	n	Sampling Fraction $\frac{n}{ND}\%$	ND	n	Sampling Fraction $\frac{n}{ND}\%$	ND	n	Sampling Fraction $\frac{n}{ND}\%$
Jan	403	6	1.49	372	5	1.34	465	5	1.08
Feb	377	8	2.12	336	4	1.19	420	4	0.95
Mar	403	7	1.74	372	6	1.61	465	4	0.86
Apr*	390	5	1.28	-	-	-	450	4	0.89
May*	403	2	0.50	-	-	-	465	3	0.65
Jun	488	6	1.23	105	2	1.90	450	6	1.33
Jul	527	6	1.14	465	6	1.29	465	6	1.29
Aug	496	4	0.81	279	4	1.43	465	5	1.08
Sep	510	5	0.98	360	5	1.39	450	3	0.67
Oct	558	5	0.90	403	3	0.74	465	4	0.86
Nov	360	5	1.39	390	5	1.28	420	5	1.19
Dec	338	4	1.18	403	3	0.74	465	3	0.65

The zone K4 consisted of 13 fish landing centres with four strata. These four strata contained 1, 2, 6 and 4 landing centres respectively. During this period, this zone contributed nearly 11% of the total landings of Kerala. Except two three months, the sampling fraction achieved were around 1%. The stratum wise coverage was not uniform over the months. In certain months only stratum 1 was observed, whereas in certain other months, only stratum 2 and 3 were observed. The landings were high during September and December.

Zone K5

Zone K5 consisted of 9 fish landing centres which were considered together without stratification. The sampling fraction was less than 2% for most of the months. High landings were noted in a few months due to the use of mechanised ring-seine units at one centre of the zone. In September 2004, only one centre was observed.

Table 2.4(5) Zonal sampling fractions during 2004-06 - Zone K5

Month	2004			2005			2006		
	ND	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$
Jan	155	2	1.29	155	4	2.58	279	2	0.72
Feb	145	2	1.38	140	3	2.14	140	2	1.43
Mar	155	2	1.29	155	2	1.29	279	2	0.72
Apr*	150	2	1.33	-	-	-	270	2	0.74
May*	150	2	1.33	-	-	-	279	2	0.72
Jun	150	5	3.33	210	3	1.43	270	4	1.48
Jul	248	5	2.02	279	3	1.08	279	5	1.79
Aug	217	4	1.84	279	2	0.72	279	2	0.72
Sep	270	1	0.37	270	3	1.11	270	2	0.74
Oct	155	2	1.29	155	2	1.29	279	2	0.72
Nov	150	2	1.33	150	2	1.33	270	2	0.74
Dec	130	2	1.54	155	2	1.29	279	2	0.72

Zone K6

There were of 19 landing centres in zone K6 and they were grouped into three strata consisting of 5, 6 and 8 centres respectively. During the period 2004-06, the zone K6 contributed nearly 13% of the total landings of Kerala. Here it was seen that all the three strata were observed in almost all the months. From the Table 2.4(6) it could be seen that the sampling fractions were around 1% in most of the months with four months registering above 2% with a maximum of 3.3% in June 2004. The landings were fluctuating over the months. In the year 2004, the maximum landings were during September-October but in the year 2005, the landings were high in December and the trend continued to January 2006.

Table 2.4(6) The sampling fraction for the zone K6 during 2004-2006

Month	2004			2005			2006		
	ND	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$
Jan	403	4	0.99	496	6	1.21	403	6	1.49
Feb	377	5	1.33	364	4	1.10	364	4	1.10
Mar	403	4	0.99	403	6	1.49	434	7	1.61
Apr*	360	4	1.11	-	-	-	390	6	1.54
May*	403	4	0.99	-	-	-	403	4	0.99
Jun	390	4	1.03	512	6	2.31	360	3	0.83
Jul	403	4	0.99	434	2	0.46	217	2	0.92
Aug	496	3	0.60	372	3	0.81	310	4	1.29
Sep	390	6	1.54	360	5	1.39	270	5	1.85
Oct	403	5	1.24	403	5	1.24	403	5	1.24
Nov	390	5	1.28	390	5	1.28	390	6	1.54
Dec	403	6	1.49	403	6	1.49	403	5	1.24

Zone K7

There were 12 fish landing centres divided into three strata in the zone. The stratum 1 contained only one centre whereas stratum 2 and 3 contained 6 and 5 centres respectively. The annual average landings of this zone accounted for nearly 12% of the total landings of Kerala during 2004-06. The sampling fraction of the zone was found to be less than 2 % in all the months except in June 2005. The representation of each stratum was not uniform. In the year 2004, the maximum landings occurred in March due to the ring-seine operations. In general, the landings were high during October-December.

Table 2.4(7) Zonal sampling fractions during 2004-06 – Zone K7

Month	2004			2005			2006		
	ND	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$
Jan	312	4	1.28	324	5	1.54	324	3	0.93
Feb	300	4	1.33	288	2	0.69	264	2	0.76
Mar	300	3	1.00	324	2	0.62	286	4	1.40
Apr*	300	3	1.00	-	-	-	182	2	1.10
May*	324	3	0.93	-	-	-	300	4	1.33
Jun	312	2	0.64	299	4	1.34	299	4	1.34
Jul	312	3	0.96	312	2	0.64	312	2	0.64
Aug	372	3	0.81	324	1	0.31	270	2	0.74
Sep	312	5	1.60	300	3	1.00	225	3	1.33
Oct	372	4	1.08	324	3	0.93	270	4	1.48
Nov	312	3	0.96	312	5	1.60	260	4	1.54
Dec	324	4	1.23	324	4	1.23	260	5	1.92

Zone K8

There were 25 fish landing centres in this zone, which were grouped into 3 strata. They contained 3, 5 and 17 centres respectively. This zone had the highest annual average landings of nearly 20% of the total landings of the state. Even though the zone accounted for one fifth of the landings, the sampling coverage was less than 3.55% in all the months. Further it was noted that all the three strata were observed in most of the months. Though the landings were fluctuating over the months, high landings were observed during March-May.

Table 2.4(8) Zonal sampling fractions during 2004-06 – Zone K8

Month	2004			2005			2006		
	ND	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$
Jan	775	8	1.03	775	10	1.29	775	8	1.03
Feb	775	8	1.03	700	6	0.86	700	6	0.86
Mar	496	10	2.02	775	4	0.52	775	8	1.03
Apr*	480	8	1.67	-	-	-	840	12	1.43
May*	434	9	2.07	-	-	-	775	9	1.16
Jun	375	9	2.40	342	12	3.51	345	12	3.48
Jul	310	10	3.23	310	4	1.29	310	7	2.26
Aug	310	11	3.55	310	5	1.61	355	9	2.54
Sep	750	8	1.07	750	6	0.80	750	9	1.20
Oct	775	10	1.29	775	5	0.65	775	8	1.03
Nov	750	10	1.33	750	5	0.67	750	9	1.20
Dec	775	11	1.42	775	9	1.16	775	8	1.03

Zone K9

There were 11 fish landing centres in this zone. The stratum 1 consists of 1 centre, stratum 2 and 3 with 5 centres each. On an average, this zone accounted for 5% of the total landings of the state during the period of study. The sampling coverage was around 1% itself for most of the months with a maximum of 2.35%. In this zone, all the strata were observed in each month. Since only one centre was observed in three strata in many months, the estimate of variance was zero for those months.

Table 2.4(9) Zonal sampling fractions during 2004-06 – Zone K9

Month	2004			2005			2006		
	ND	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$
Jan	403	5	1.24	403	9	2.23	341	8	2.35
Feb	377	5	1.33	336	6	1.79	308	3	0.97
Mar	403	5	1.24	372	5	1.34	341	2	0.59
Apr*	390	5	1.28	-	-	-	330	3	0.91
May*	403	4	0.99	-	-	-	341	3	0.88
Jun	390	4	1.03	344	3	0.87	315	4	1.27
Jul	372	6	1.61	372	3	0.81	341	4	1.17
Aug	372	3	0.81	372	2	0.54	341	3	0.88
Sep	390	8	2.05	270	3	1.11	330	4	1.21
Oct	403	6	1.49	341	2	0.59	341	3	0.88
Nov	390	5	1.28	330	3	0.91	330	3	0.91
Dec	403	7	1.74	341	4	1.17	341	3	0.88

Zone K10

There were 17 fish landing centres in this zone being divided into three strata with 2, 5 and 10 centres respectively. The annual average landings was about 4% of the total landings of the state. The sampling coverage was less than 1.5% in all the months. The estimates of landings were fluctuating over the months and over the years.

Table 2.4(10) Zonal sampling fractions during 2004-06 – Zone K10

Month	2004			2005			2006		
	ND	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$	*ND*	*n*	Sampling Fraction $\frac{n}{ND}\%$
Jan	584	6	1.03	558	8	1.43	527	4	0.76
Feb	522	5	0.96	504	7	1.39	476	4	0.84
Mar	558	5	0.90	558	6	1.08	527	4	0.76
Apr*	540	4	0.74	-	-	-	510	4	0.78
May*	558	5	0.90	-	-	-	527	6	1.14
Jun	270	1	0.37	-	-	-	255	1	0.39
Jul	558	2	0.36	465	1	0.22	341	5	1.47
Aug	558	3	0.54	558	1	0.18	527	3	0.57
Sep	540	6	1.11	510	2	0.39	510	2	0.39
Oct	558	6	1.08	527	4	0.76	-	-	-
Nov	540	8	1.48	510	3	0.59	510	5	0.98
Dec	558	8	1.43	527	4	0.76	527	4	0.76

Single Centre Zones

The five single centre zones had an average contribution of nearly 9% at Neendakara, 6% at Sakthikulangara, 4% at Vypin and 3% each at Cochin fisheries harbour and Munambam. The sampling fraction achieved in the single centre zones were on an average 20%. This is due to assigning one field staff exclusively for each of the zones.

Table 2.4(11) Zonal sampling fractions during 2004-06 – Neendakara

Month	2004			2005			2006		
	N	n	Sampling Fraction $\frac{n}{N}\%$	N	n	Sampling Fraction $\frac{n}{N}\%$	N	n	Sampling Fraction $\frac{n}{N}\%$
Jan	27	6	22	26	4	15	26	8	31
Feb	24	6	25	24	3	13	24	7	29
Mar	27	3	11	27	6	22	27	7	26
Apr*	26	2	8	-	-	-	25	8	32
May*	26	3	12	-	-	-	27	9	33
Jun	12	3	25	12	4	33	12	3	25
Aug	26	7	27	27	7	26	28	6	21
Sep	26	7	27	26	4	15	26	8	31
Oct	26	5	19	26	5	19	26	7	27
Nov	26	6	23	26	7	27	26	4	15
Dec	12	3	25	27	7	26	26	6	23

Table 2.4(12) Zonal sampling fractions during 2004-06 – Sakthikulangara

Month	2004			2005			2006		
	N	n	Sampling Fraction $\frac{n}{N}\%$	N	n	Sampling Fraction $\frac{n}{N}\%$	N	n	Sampling Fraction $\frac{n}{N}\%$
Jan	27	6	22	26	4	15	26	6	23
Feb	24	4	17	24	2	8	24	9	38
Mar	27	6	22	27	6	22	27	8	30
Apr*	26	3	12	-	-	-	25	7	28
May*	26	3	12	-	-	-	27	6	22
Jun	12	4	33	12	4	33	12	4	33
Aug	26	7	27	27	7	26	28	6	21
Sep	26	6	23	26	5	19	26	8	31
Oct	26	6	23	26	6	23	26	7	27
Nov	26	4	15	26	6	23	26	7	27
Dec	12	2	17	27	7	26	26	7	27

Table 2.4(13) Zonal sampling fractions during 2004-06 – Cochin fisheries harbour

Month	2004			2005			2006		
	N	n	Sampling Fraction $\frac{n}{N}\%$	N	n	Sampling Fraction $\frac{n}{N}\%$	N	n	Sampling Fraction $\frac{n}{N}\%$
Jan	27	6	22	26	5	19	26	4	15
Feb	24	6	25	24	3	13	24	3	13
Mar	27	4	15	27	3	11	27	4	15
Apr*	26	5	19	-	-	-	25	3	12
May*	26	5	19	-	-	-	27	3	11
Jun	12	3	25	12	3	25	12	3	25
Jul	27	3	11	26	2	8	26	3	12
Aug	26	4	15	27	3	11	26	4	15
Sep	26	5	19	26	3	12	26	2	8
Oct	26	4	15	26	4	15	26	3	12
Nov	26	4	15	26	3	12	26	4	15
Dec	14	3	21	27	3	11	26	4	15

Table 2.4(14) Zonal sampling fractions during 2004-06 – Munambam

Month	2004			2005			2006		
	N	n	Sampling Fraction $\frac{n}{N}\%$	N	n	Sampling Fraction $\frac{n}{N}\%$	N	n	Sampling Fraction $\frac{n}{N}\%$
Jan	31	5	16	31	4	13	31	3	10
Feb	29	5	17	28	5	18	28	1	4
Mar	31	5	16	31	2	6	31	3	10
Apr*	30	7	23	-	-	-	30	3	10
May*	30	4	13	-	-	-	31	3	10
Jun	14	3	21	14	1	7	14	2	14
Jul	31	1	3	31	2	6	31	2	6
Aug	31	2	6	31	2	6	32	3	9
Sep	30	3	10	30	3	10	30	1	3
Oct	31	2	6	31	3	10	31	1	3
Nov	30	3	10	30	2	7	30	1	3
Dec	17	2	12	31	3	10	31	2	6

Table 2.4(15) Zonal sampling fractions during 2004-06 - Vypin

Month	2004			2005			2006		
	N	n	Sampling Fraction $\frac{n}{N}\%$	N	n	Sampling Fraction $\frac{n}{N}\%$	N	n	Sampling Fraction $\frac{n}{N}\%$
Jan	27	3	11	26	4	15	26	4	15
Feb	24	4	17	24	4	17	24	3	13
Mar	27	4	15	27	2	7	27	4	15
Apr*	26	3	12	-	-	-	25	4	16
May*	26	3	12	-	-	-	27	2	7
Jun	12	2	17	12	2	17	12	2	17
Jul	27	1	4	26	3	12	13	2	15
Aug	26	1	4	27	2	7	14	3	21
Sep	26	3	12	26	3	12	26	3	12
Oct	26	2	8	26	3	12	26	4	15
Nov	26	2	8	26	4	15	26	3	12
Dec	14	2	14	27	3	11	26	4	15

The tables 2.4 and the discussions in general indicate the following:

(i) The sampling fractions achieved in the multi-centre zones were very small.

(ii) Very poor coverage of the multi-centre zones. In cases where the zone is divided into 3 or 4 strata, the survey was often concentrated on one or two strata only.

(iii) Though 9 days of 24 hour duration are proposed for observation in each zone, very often only 4 or 6 days are achieved.

(iv) When the number of days is too less, in zones having more than one stratum, very often either one or two strata get excluded or may be represented only with one observation. In either case an estimate of the stratum variability cannot be obtained. This undermines the very purpose of stratification and variance estimation. As a result the estimate of zonal variance fails to represent anything somewhere close to real situation.

(v) The fish landings data is highly unstable. Even the minimum and maximum catches are also distributed over the months differently over the zones and years.

Some of the limitations of the existing sampling scheme are described in the section below.

2.4 Limitations of the Existing Design

The Fast Changing Fisheries sector and a fixed sampling design

The currently followed sampling design was developed decades back. When the design was developed, the boats used for fishing were non-mechanised, fishing operations were limited to single day and only one or two types of gears were employed. At present, most of the boats operating are either motorised or mechanised, instead of single gear, multi gears are used, providing wider coverage and efficient catchability. Similarly, the increase in the time spent for fishing in the mechanised sector by undertaking multiple days of fishing, use of sophisticated electronic devices for identifying the fish species and the locations of their concentration and communications has resulted in increased fishing pressure and fishing efficiency. Though there were changes and improvements from time to time in view of practical contingencies in the sampling design, the basic structure of the methodology are same as developed in fifties. One of the main changes has been in the selection process of space- time units i.e., the landing centre days (first stage units). Earlier, a landing centre was observed continuously for a number of days, then the field staff were to move to the next randomly selected landing centre which also was observed continuously for a few days. It was similar to a two dimensional population with landing centres over one dimension and days over the other dimension. Later, the landing centre day formed the primary stage unit, reducing a two dimensional structure into a one dimensional structure. As a result, a centre selected to the sample may or may not be observed for more than one day. Further, the selection of second stage units does not give any weightage with respect to craft or gear operated.

Low Sampling fraction

One of the main limitation in the existing sample survey is the use of a very low sampling fraction. From the tables 2.4(1) to (10) corresponding to the multi-centre zones the sampling fraction achieved is around 1% barring a few months which register a value around 2 or 3%. In the case of single centre zones, tables 2.4(11) to (15), the sampling fraction achieved was on an average between 15 to 20%. The existing methodology is to select 9 landing centre days at random for each zone irrespective of the total number of landing centre days in it and assign them to the landing centres stratum wise for multi-centre zones. In this assignment quite often one or more of the strata may not get any representation in the sample at all.

To get an idea about the minimum sample size required to estimate the total landings in a zone, the sample sizes are estimated using the formula (2.24) in the case of single centre zones. The results for the three single centre zones based on the data for the year 2006 are given in table 2.5. T indicates that the minimum sample size required to get the estimate within 10% margin of error is above 50%. While for within 20% margin of error, it is around 20 to 30%. Note that the requirement of very high sample sizes is due to the existence of very high variability in the data compared to the mean. This reveals that the sampling fraction achieved at present in the multi-centre zones are alarmingly low and those achieved in single centre zones are also far below the required level. The major drawback in depending on a very low sample size in a highly variable population is that the inferences will be totally misleading.

Table 2.5 Estimated sample sizes for specified margin of error for the year 2006

Error (%)	Days	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Neendakara*													
	N	26	24	27	25	27	12	-	28	26	26	26	26
10	n	15	14	18	15	16	12	-	6	12	21	23	21
15		10	10	12	10	10	11	-	3	7	17	21	16
20		7	7	9	7	7	11	-	2	4	13	18	13
Sakthikulangara*													
	N	26	24	27	25	27	12	-	28	26	26	26	26
10	n	17	16	14	13	23	11	-	24	10	13	14	16
15		12	12	8	8	19	10	-	20	6	8	9	11
20		9	8	5	5	15	9	-	16	4	5	6	8
Cochin													
	N	26	24	27	25	27	12	26	26	26	26	26	26
10	n	19	22	13	22	20	12	23	20	9	24	16	24
15		15	20	8	19	16	11	19	15	5	21	11	21
20		11	18	5	16	12	11	16	12	3	19	8	19

*No landings during the month of July due to ban on fishing

Period of observation

The division of the 24 hours duration of an observation day into three independent seems to be inappropriate. One drawback of this is due to the failure to get any reliable details regarding the night landings. Since a lot of facilities are available for moving the fish landed to any distant location within hours and in the absence of any kind of recording of the boats landed, there is very little scope to get the required details of night landings, unless the field staff remain at the centre during night. The present practice of collecting the details of the night landings by enquiry at the next day morning may provide at most the total count of the boats landed.

Lack of sampling frame for SSUs

A fishing boat is selected as the second stage unit. One of the main difficulties in the selection of fishing boats is the lack of sampling frame. There are fishing boats operating without registration. A complication is that it is difficult to know in advance when, where and how many landings will occur. The timing of landings might be influenced by season, weather conditions, day of the week, fishing location and a host of other potential factors.

In the existing sampling design, when the total number of fishing boats landed is 15 or less, the landings from all the units are enumerated. When the total number of units exceeds 15, the sampling is done in a predetermined manner given in section 2.2.3. At present, for example, the observation of the landings by a mechainsed trawler unit will take at least 20 minutes time, due to the wide variety of fish species they often carry. Though the time taken for observing a fishing boat varies depending upon the type of craft and gear of the unit, in practical situation, it may not be possible to cover more than 20 fishing boats (approximately 20 minutes for each unit) within a six hour period of observation.

The methodology enumerates the landings from the total fishing boats landed. At present different types of fishing boats are operating in a landing centre. The systematic selection of the boats will give a random sample of the fishing boats, but it may not be possible to give representation to each type of boat that lands in a period of time. The estimate will be more precise if different classes of boats (either in terms of size; large, medium, small, or of type of gear used; lines, gill net, ring seine or mechanised, motorised, artisanal etc.) are sampled and analysed separately.

Gear-wise recording of landings

Since the number of boats operating has increased tremendously, it is very difficult to select the sample from the total boats by the existing sampling scheme. Also, it is observed that fishermen use one type of fishing gear during one fishing season and a different one during another season. It is also found that fishing boats use two or more gears simultaneously. In these circumstances, it is not possible to estimate the proportion of catch that has resulted from each gear separately, unless the different gears are targeting completely different species. Most of the times, the predominant gear is used to describe the boat/gear type.

Estimation of total landings

In the selection procedure of PSU's , first 9 clusters of two days were selected in a systematic manner and then 9 centres were allotted to the selected 9 cluster days, where the centres were selected by simple random sampling with replacement from the total number of landing centers in a zone. But, in the estimation procedure, the PSU's are considered as samples selected by simple random sampling without replacement. That is, the two dimensional character of the fish landings is disregarded. In that case, when the variance of the fish landings is estimated, it is not appropriate to consider the PSU's as having been drawn at random from all landing centre–days in a zone. Furthermore, the variance of the fish landings will depend on the number of landing centres and the number of days per landing centre that have been chosen.

Further, from the estimation procedure described in the previous section, it could be seen that Table 2.3 was made by modifying each cell of the Table 2.2 by replacing the observed landings with the estimates of landings for each type of gear during the p^{th} period of observation. Further, the month totals are estimated by taking gear-wise average on each day and then combining those averages. Altogether this technique has less statistical validity.

Estimation of variance

The primary concern in all sample surveys is the derivation of point estimates for the parameters of main interest. However, equally important is the derivation of their variances. The estimated variance is a main component of the quality of any estimator. According to Gagnon et al. (1997), the variance estimation provides a measure of the quality of estimates; is required for the computation of confidence intervals; helps to draw accurate conclusions and allows statistical agencies to give users indications of data quality. Thus, variance estimation, understanding and reducing undesirable variation in the estimate are crucial issues in the assessment of the survey results.

In the earlier fishery survey, Sukhatme et al. (1958) have shown that the correlation between the number of boats landing per hour and the average catch per boat for the same hour is small. They have further shown that the coefficient of variation of the catch per boat is also small. Hence the error introduced by sampling the second stage units remains minimal and can therefore be ignored. Thus the variance is estimated as the variance between days, assuming that variance between boats of the same gear to be negligible. This is mainly because of the indigenous method of fishing in which the fishing power, the gears employed, the time spent on fishing and the area fished remain nearly same. The present method of computing the variance of the total landings by using the formula (2.14) takes into account the variation due to days only. Due to the use of high capacity boats with multi-gears and different varieties of nets and other fish catching devices, there is every possibility to expect appreciable variations due to both boats and gears. So the present method of accounting only for variations due to days is inadequate to expose the true variability in the data.

2.5 Remedial Measures

The major limitations among those listed in the previous section are

The low sampling fraction and

The method of variance estimation

A sampling scheme is valid if and only if the sample selected is a typical representative part of the population. This characteristic of the sample can be ensured only if, among other things, the sample size also is increased along with increase in the heterogeneity in the population. Hence every possible effort has to be made to increase the sample size, particularly in the multi-centre zones.

As far as the estimation of the variance is concerned, efforts have to be made to expose the different sources of variation in the fishery data. Once the sources of variations and their impact on the estimate are identified, suitable methods have to be adopted to extract the true variances due to each of the sources and utilize them in the decision making. The rest of the limitations can be rectified without much effort.

The procedure of dividing a day into three periods and considering each of them independently may be discontinued by treating the day as a single unit. Both the count and catch may be recorded simultaneously during day times and the night landings may be used for count only. By taking the 24 hour duration of a day as a single unit, we get more opportunity to ensure adequate representation to all different kinds of boats that land during the day. The selection of boats for recording catch may be made in such a way that all suspected cases of heterogeneity may get adequate representation in the sample. Methods of adjusting the estimates may be avoided to the extent possible. In the remaining chapters of the thesis we shall examine how far these suggested remedial measures can be incorporated in the design and estimation procedures.

CHAPTER III

Marine Fish Landings in Kerala – Fluctuations and Trends

3.1 Introduction

In this chapter an attempt is made to assess the fluctuations and trends in the data on fish landings, the number of boats operated and landings/boat recorded so far. Our objective here is to draw an idea about the overall variability in the data and exploit the trends and patterns if any for predictions. Section 3.2 describes how the fisher folk population and their fishing activity varied over time. In section 3.3 the trends in the total assemblage wise and sector wise landings are described by fitting suitable trends to the data. In section 3.4 the trends in the landings, boats operated and the landings per boat are discussed in detail by fitting trends for monthly data based on 2 term, 3 term, 4 term and 6 term moving averages. Comparisons of the period wise landings at the five single centre zones are made in section 3.5. Section 3.6 concludes the findings in the analysis made in the chapter.

3.2 Changes in the Fisher-Folk Population, Craft and Gears

State wide census on all characteristics of marine fishery was conducted in 1980 and 2005. The number of fishing villages, number of fishermen families and fisher folk population in Kerala during 1980 and 2005 census are given in Table 3.1. The number of persons actually involved in fishing activity is given in Table 3.2

Table 3.1 Fisher folk population 1980-2005

District	Number of fishing villages		Number of fishermen families		Fisher folk population	
	1980	2005	1980	2005	1980	2005
Thiruvana-nthapuram	54	42	26,519	34,128	132,087	143,436
Kollam	29	26	12,381	11,899	79,113	43,210
Alappuzha	39	30	15,648	21,759	97,388	101,341
Ernakulam	20	21	7,648	8,876	49,059	42,069
Thrissur	22	18	8,295	6,598	60,432	34,078
Malappuram	18	23	8,321	10,462	70,904	79,858
Kozhikode	57	35	11,884	16,058	79,434	87,690
Kannur	65	11	9,148	5,929	71,455	36,686
Kasaragod *		16	-	4,777	-	33,866
Total	304	222	99844	120,486	639,872	602,234

Source: CMFRI (1981) & DAHD&F(2006)

* In the 1980 census Kasaragod was included in Kannur itself

In the year 2005, the number of fishing villages was decreased by 27% as compared to the year 1980. The number of fishermen families during 2005 registered an increase of 21% compared to that of 1980. The fisher folk population showed a slight reduction of 6% during 2005.

Table 3.2 Fisher folk population involved in fishing as per census 1980 and 2005

District	Full time		Part time		Occasional		Total	
	1980	2005	1980	2005	1980	2005	1980	2005
Thiruvana-nthapuram	20,882	32,199	5,115	4,586	3,116	2,020	29,113	47,036
Kollam	12,115	8,255	875	201	982	209	13,972	10,522
Alappuzha	19,365	23,783	904	1,079	872	393	21,141	27,031
Ernakulam	7768	7,707	1,862	1,638	586	368	10,216	12,161
Thrissur	10,186	6,329	720	261	700	464	11,606	8,474
Malappu-ram	12,944	14,384	425	992	577	1,046	13,946	17,424
Kozhikode	16,005	18,740	435	751	609	628	17,049	21,163
Kannur	12,705	5,837	681	332	672	301	14,058	7,823
Kasaragod	-	6,869	-	648	-	202	-	7,719
Total	111,970	124,103	11,017	10,488	8,114	5,631	131,101	159,353

Source: CMFRI (1981) & DAHD&F(2006)

The total marine fishermen population engaged in actual fishing showed an increase of 22% in 2005 compared to 1980. Though there is an increase of 11% in the fishermen population engaged in full time fishing, the part time and occasional showed a decrease of 5 and 31% respectively. The fishing craft operated are broadly classified into three sectors – *mechanised, motorised* and *non-mechanised*. The sector wise craft and boats as per the 1998 and 2005 census are given in Table 3.3. (The data on the different boats, craft for the state as a whole were not available in the 1980 census hence the data collected in the rapid census in 1998 is used). The total number of craft showed a decrease of 4% in 2005 as compared to 1998.

Table 3.3 Number of craft in the fishery during 1998-2005

District	Mechanised		Motorised		Non-mechanised		Total	
	1998	2005	1998	2005	1998	2005	1998	2005
Thiruvana-nthapuram	24	55	2705	3063	6560	5005	9289	8123
Kollam	1577	1272	729	605	943	425	3249	2302
Alappuzha	0	136	4196	3947	697	1010	4893	5093
Ernakulam	1989	1898	254	1104	402	1190	2645	4192
Thrissur	60	259	538	456	211	306	809	1021
Malappuram	247	441	1903	1607	635	361	2785	2409
Kozhikode	746	1034	2375	1976	676	641	3797	3651
Kannur	239	226	868	503	402	290	1509	1019
Kasaragod	206	183	1094	890	195	294	1495	1367
Total	5088	5504	14662	14151	10721	9522	30471	29177

Source: CMFRI (1981) & DAHD&F(2006)

The variations in the total fish landings over the past years and their patterns are described in the following section.

3.3 Trends in the Total Marine Fish Landings

The estimates of marine fish landings made by CMFRI were used for the trend analysis. The introduction of new technologies and changed mode of fishing operations over the years had raised the state's marine fish production from about 2.68 lakh tonnes in the year 1961 to 5.92 lakh tonnes in 2006 with an all time peak production 6.63 lakh tonnes during 1990. The annual landings in Kerala have shown very high fluctuations over the years (Figure 3.1(a)). The average catch per year during 1961-70, 1971-80, 1981-90, 1991-2000 and 2001-06 were 3.06, 3.68, 4.17, 5.67 and 5.77 lakh tonnes respectively. On the whole there was increase in the marine fish landings in Kerala. The introduction of outboard motors to the traditional craft, in the early eighties made conspicuous impact on the marine fisheries sector in Kerala (Balan *et al.*, 1989). There was a spurt in the total landings recording 4.69 lakh tonnes during 1989 and in 1990, the landings recorded an all time high of 6.63 lakh tonnes. These increase in the landings were mainly due to bumper catches by the ring seines. However, the increase could not maintain for long. In 1991, the total landings decreased to 5.64 lakh tonnes and there after remained more or less steady.

Polynomials of best fit.

over the period 1961-2006,

$$y = 7E\text{-}07x^5 - 0.0001x^4 + 0.005x^3 - 0.1057x^2 + 0.9293x + 0.8625, \quad (3.1)$$

over the period 1971-2006,

$$y = -0.0005x^3 + 0.0288x^2 - 0.3601x + 4.6711, \quad (3.2)$$

over the period 1981-2006,

$$y = 0.0003x^3 - 0.023x^2 + 0.5127x + 2.0936, \quad (3.3)$$

over the period 1991-2006,

$$y = 0.0192x + 5.5527, \quad (3.4)$$

over the period 2001-2006,

$$y = 0.0729x^3 - 0.8652x^2 + 3.0708x + 2.7992. \quad (3.5)$$

In an attempt to find the polynomial of best fit to the fish landing data from 1961 to 2006, we got the polynomial of 5th degree given by equation (3.1). The search for trends was repeated by discarding the data for the successive decades starting from 1961. The curves of best fit obtained are given by equation (3.2) to (3.5). Figures 3.1(a) to 3.1(e) exhibit the data with the curves of best fit superimposed on each of them respectively. Note that while the trend over 45 years (1961-2006) is in the form of a 5th degree polynomial, those over lesser time periods are in the form of 3rd degree polynomial except over the period 1991-2006 which is linear. The straight line trend over 1991 to 2006 is a reflection of the stability in the data over the decade 1991-2001. It is interesting to see that when the stable decade 1991-2001 is excluded, the trend again assumed the form of a 3rd degree polynomial over the remaining period 2001-2006. Though 3rd degree polynomials were seen to be the curve of the best fit, only the degree of the polynomial remains the same while the coefficients of the like terms differ considerably in their magnitude and sign. This indicates that it is not possible to describe the fish landings using a mathematical function over any future period, however short the period may be. The analysis reveals that the fish landing data are highly unstable and unpredictable.

Figure 3.1 Trend in total marine fish landings in Kerala during 1961-2006

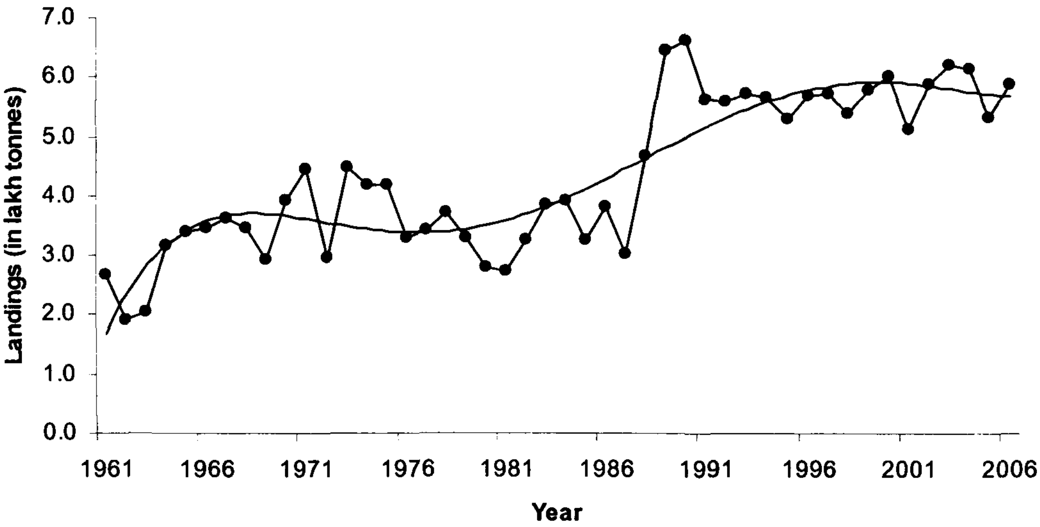


Figure 3.1(a)

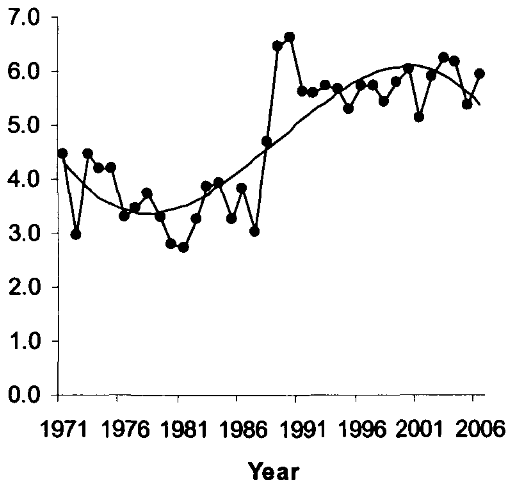


Figure 3.1(b)

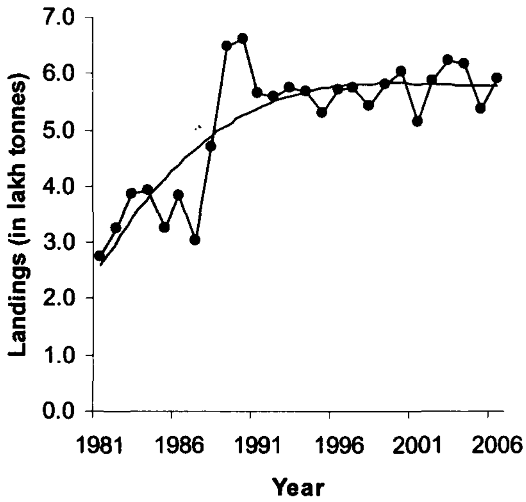


Figure 3.1(c)

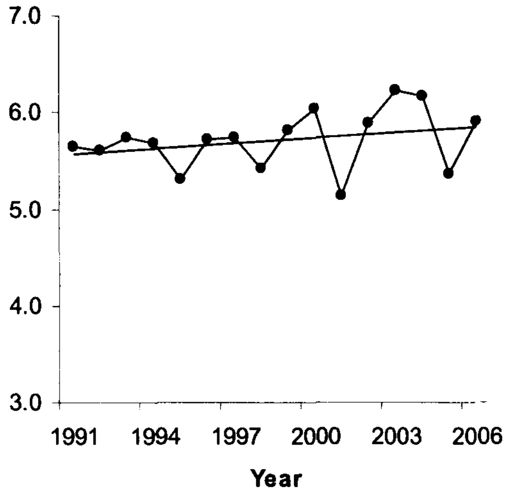


Figure 3.1(d)

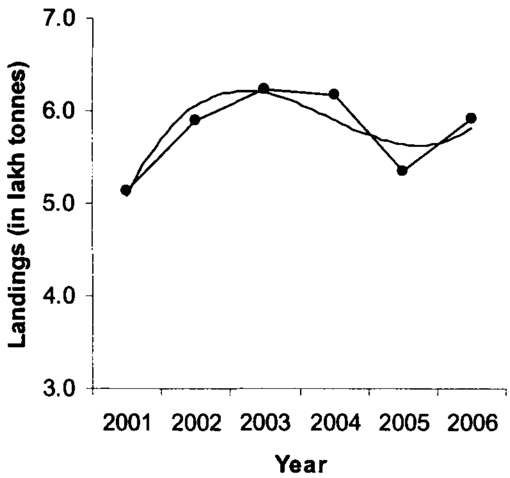


Figure 3.1(e)

Assemblage-wise landings

The total landings are often divided into two major groups, namely the pelagic and demersal. The free resources living in the water masses specifically, at the surface and subsurface waters are called pelagic resources and those found at the benthic realm are called demersal resources. The pelagic group comprises fishes such as oil sardine, lesser sardines, *chirocentrus, hilsa shad,* other shads, *stolephorus, thryssa, setipinna, coilia,* other clupeids, Bombayduck, half-beaks, full-beaks, flyingfish, ribbonfish, carangids, mackerel, seerfish, tunas, barracudas and mullets. The Demersal group comprises fishes such as elasmobranches, eels, catfishes, lizardfishes, red mullets, polynemids, sciaenids, silverbellies, lactarius, pomfrets, soles, prawns, lobsters and cephalopods.

The main characteristic of the marine fish landings in Kerala is the predominance of the pelagic resources. In the year 2006, the annual pelagic fish landings are estimated at 4.19 lakh tonnes and the demersal

1.71 lakh tonnes, the former accounting for about 71% and the latter 29%. It is also seen that demersal fin fishes account for about 14%, crustaceans 10% and molluscs 5% of the total landings. Curves of best fit are tried for the assemblage wise landings as in the case of total landings. The behaviour was still worse in the case. Over the 45 year period, the cuves of best fit were of the 6th degree given by equation (3.7) and (3.8) for the pelagic and demersal groups respectively. Figure 3.2 exhibits the assemblage wise data with the polynomials of best fit superimposed on it.

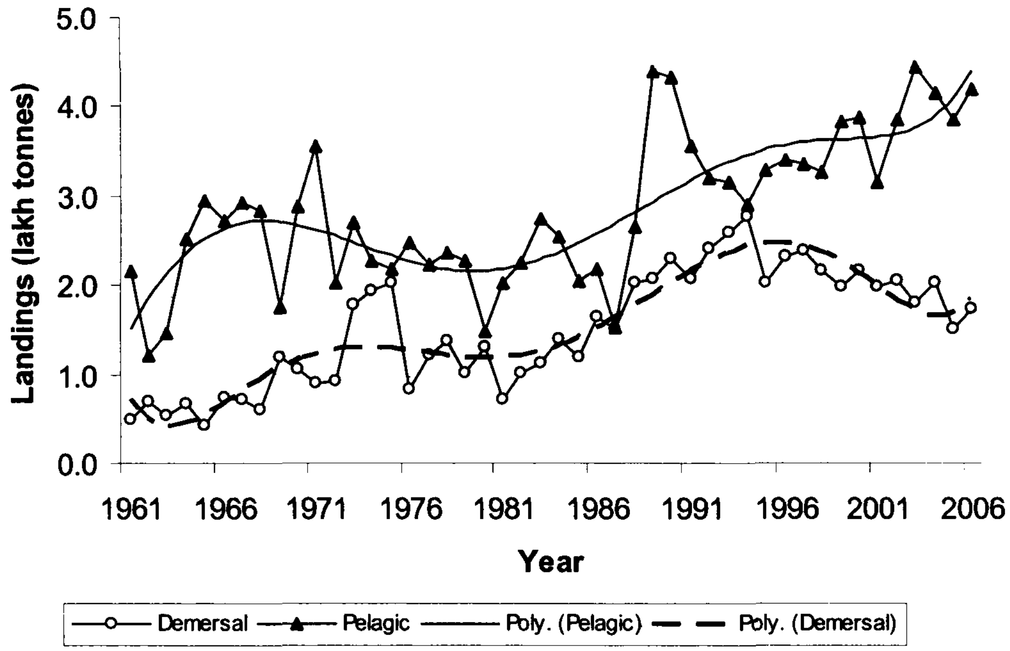
The curve of best fit for the pelagic landings, over the period 1961 – 2006,

$$2E\text{-}08x^6 - 2E\text{-}06x^5 + 8E\text{-}05x^4 - 5E\text{-}05x^3 - 0.0346x^2 + 0.4525x + 1.0911 \quad ...(3.7)$$

The curve of best fit for the demersal landings, over the period 1961 – 2006,

$$6E\text{-}08x^6 - 9E\text{-}06x^5 + 0.0005x^4 - 0.0114x^3 + 0.1281x^2 - 0.5203x + 1.1188 \quad ...(3.8)$$

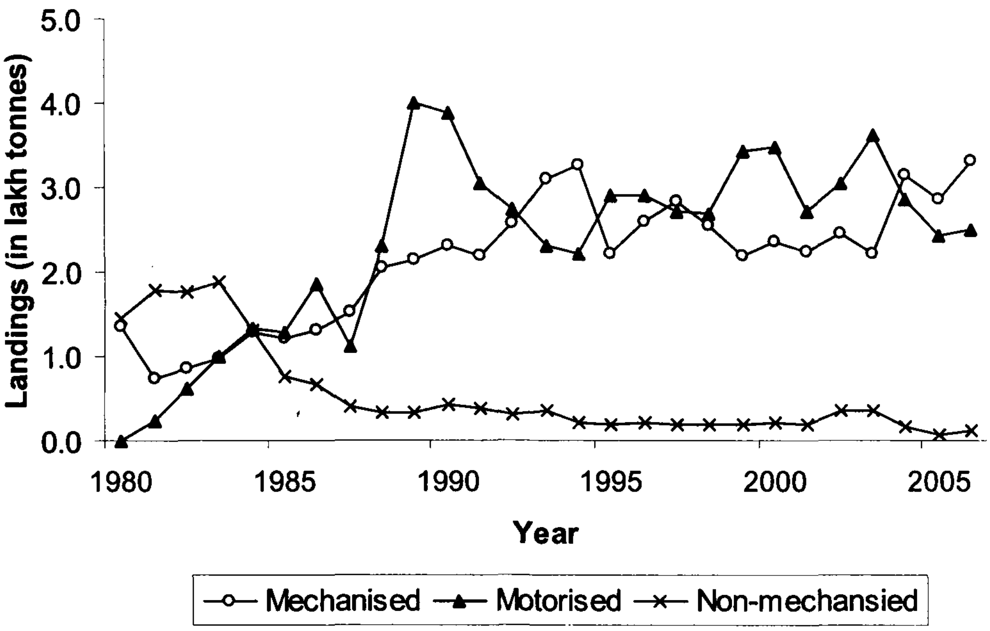
Figure 3.2 Pelagic and demersal fish landings during 1961-2006



Sector-wise landings

The sector-wise data were available from 1980 onwards only. During the year 1980 around 52% of the marine fish landed in Kerala was by the non-mechanised craft. After the motorisation of the country-craft, the contribution from the non-mechanised country-craft which was 65% in 1981, came down to 2% in 2006. The contribution from mechanised sector varied from 25% in 1983 to 56% in 2006. On the other hand the contribution from the motorised country-craft increased from 8% in 1981 to 62% in 1989. The contribution from this sector in the year 2006 is around 42%. The shift from non-mechanised fishing to mechanised and motorised fishing can be seen from Figure 3.3

Figure 3.3 Contribution of different sectors to the marine fish landings



3.4 Zone level Trends

The estimates of marine fish landings, the fishing effort expended in terms of units operated and landings per boat for each zone from 2002 to 2006 were used to examine the variability and trend in the fisheries. The landings by the three sectors, viz., mechanised, motorised and non-mechanised were plotted against the months over the years to see the trends if any in the landings. Moving averages (MA) of order 2, 3, 4 and 6 are calculated to eliminate or minimize the fluctuations in the landings data so that any underlying trend can be recognized. Similarly for the data on fishing boats and landings per boat were made. Due to space constraints only selected graphs corresponding to four representative zones, K1, K6, K8 and K11 (Neendakara) are included. Three separate graphs; one for total landings and their 3 and 6 term moving averages, the second for the number of boats landed and their moving averages and the third for the landings per boats are given.

Zone K1

In this zone only two types of boats motorised and non-mechanised which accounted for nearly 86% and 14% of the landings were in use. Figure 3.4(a) to (f) represents the trends in the total landings and their moving averages. While the landings by motorised boats were highly flexible, those by non-mechanised boats exhibited high flexibility in 2002 and 2006. The moving average plots indicate that the total landings by the motorised boats were very high during 2002 and 2006 while those due to non-mechanised were high only in 2002. In general, the landings by motorised boats increased in the later half of every year, however the landings by non-mechanised sector indicated slight decrease during the years except 2002 and 2006. Figure 3.5(a) to (d) represent the trends in the number of boats landed during the period in the two sectors. These figures do not exhibit any pattern other than the number of motorised boats landed during 2002 were comparatively high. Figure 3.6(a) and (b) representing the landings per boat indicate high variability in case of motorised boats while those in the non-mechanised category, high variability were noted only during 2002 and 2006. In general it may be concluded that neither the total landings, boats operated or the landings per boat exhibited any recognisable pattern other than the total landings by motorised sector is slightly more during the second half of every year.

Figure 3.4 Trend in the landings at zone K1 during 2002-06

Motorised

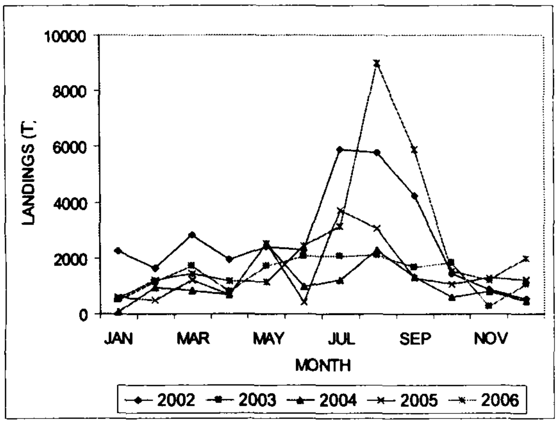


Figure 3.4(a)

Non-mechanised

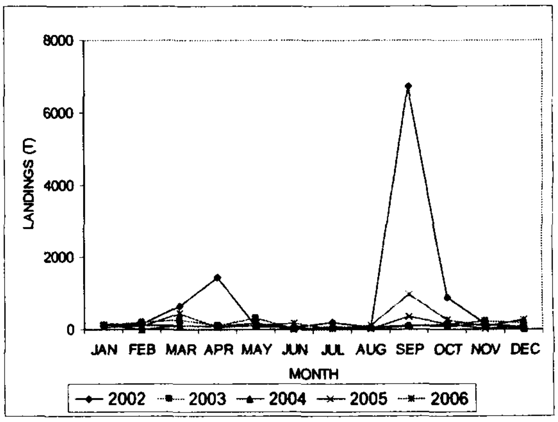


Figure 3.4(b)

Moving Average of order 3 of the fish landing series

Motorised | **Non-mechanised**

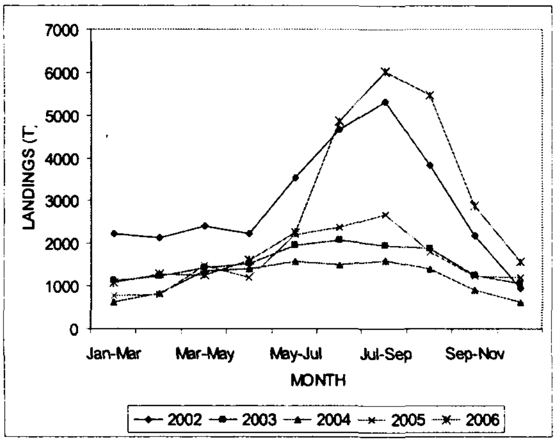


Figure 3.4(c)

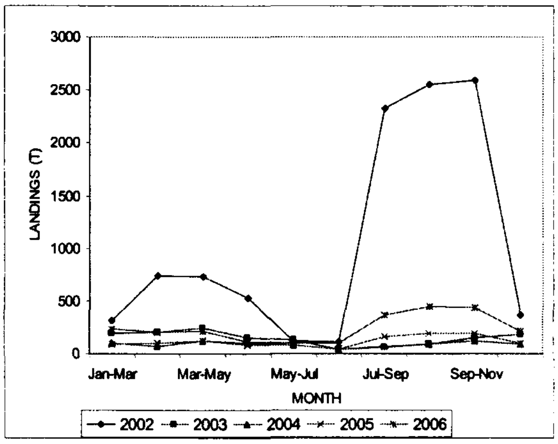


Figure 3.4(d)

Moving Average of order 6 of the fish landing series

Motorised | **Non-mechanised**

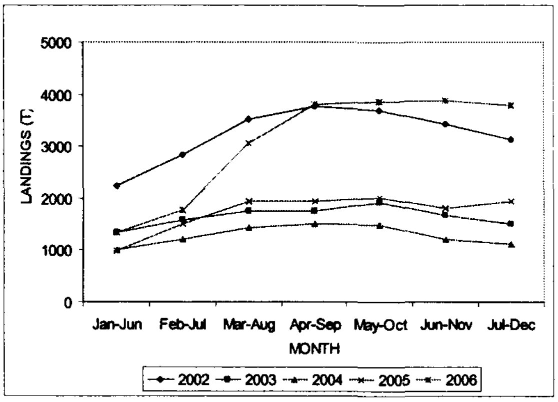


Figure 3.4(e)

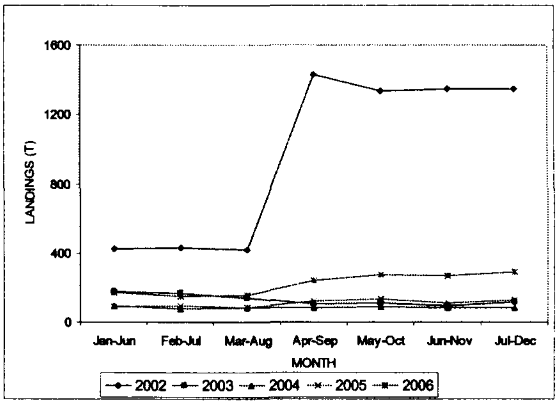


Figure 3.4(f)

Figure 3.5 Trend in the fishing boats operated at zone K1 during 2002-06

Motorised | **Non-mechanised**

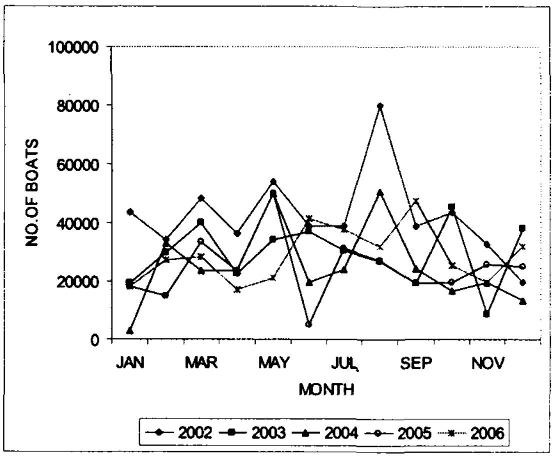


Figure 3.5(a)

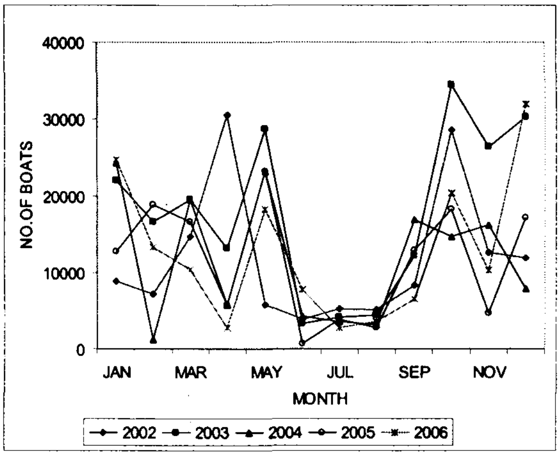


Figure 3.5(b)

Moving Average of order 3 of the fishing boats series

Motorised

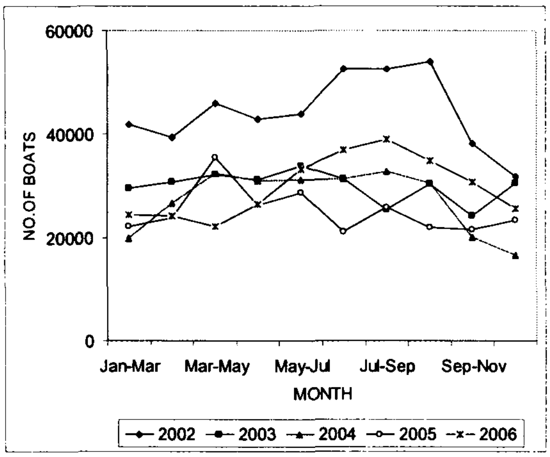


Figure 3.5(c)

Non-mechanised

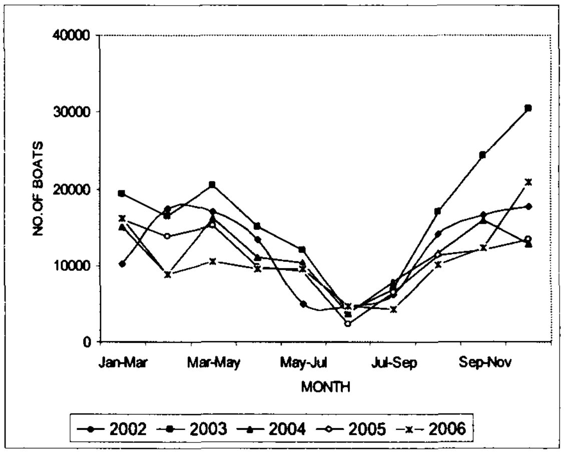


Figure 3.5(d)

Figure 3.6 Trend in the landing per boat at zone K1 during 2002-06

Motorised

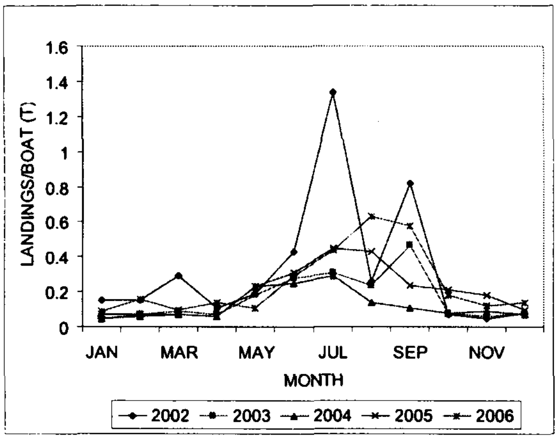


Figure 3.6(a)

Non-mechanised

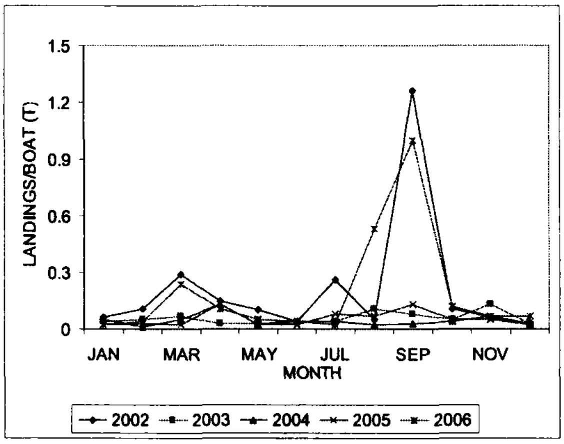


Figure 3.6(b)

Zone K6

In this zone boats of all the three categories were in use, contributing nearly 59% by mechanised, 40% by motorised, and 1% by non-mechanised sector. Figure 3.7(a) to (i) represent the behaviour of the category wise total landings and their 3 term and 6 term moving averages. The total landings are quite irregular except in the non-mechanised sector. In the non-mechanised sector high fluctuations were observed only during 2003 and 2006. The moving average graphs intended to smoothen the fluctuations exhibit a clear increasing trend over the months every year in the total landings by the motorised sector, the trend in the mechanised sector is somewhat increasing and those in the non-mechanised sector is irregular. Figure 3.8(a) to (f) for the number of boats landed and their 3 term moving averages indicate no pattern other than operation of comparatively more number of mechanised bots during January-March period every year. Figure 3.9(a) to (c) exhibiting the behaviour of landings per boat also indicate that the landings per boat also are quite irregular in all the three categories except that the landings per boat for the non-mechanised category are relatively small during the first half of the year.

Figure 3.7 Trend in the landings at zone K6 during 2002-06

Mechanised **Motorised** **Non-mechanised**

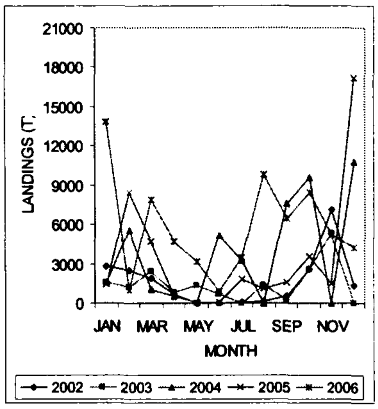


Figure 3.7(a)

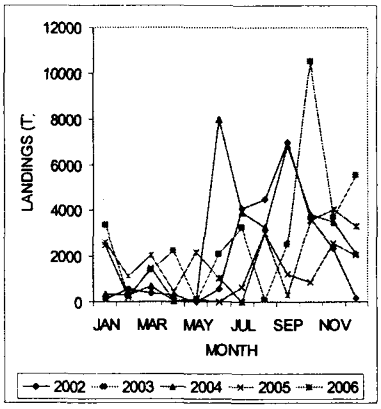


Figure 3.7(b)

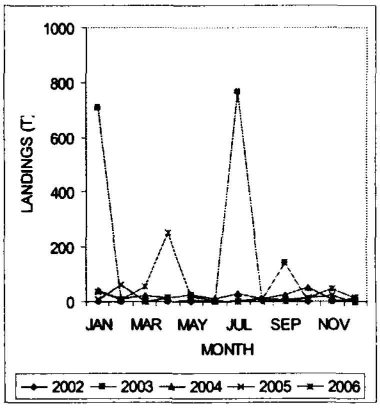


Figure 3.7(c)

Moving Average of order 3 of the fish landing series

Mechanised

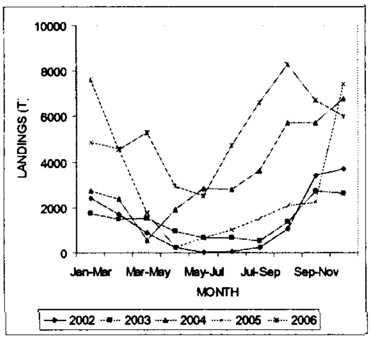


Figure 3.7(d)

Motorised

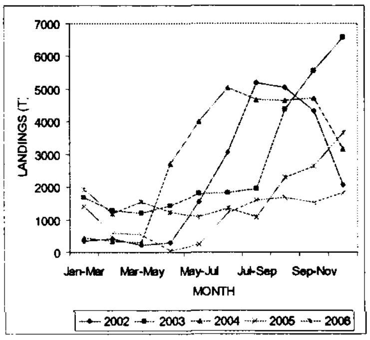


Figure 3.7(e)

Non-mechanised

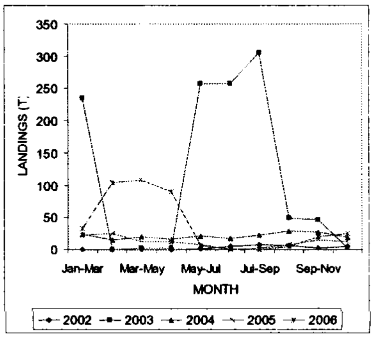


Figure 3.7(f)

Moving Average of order 6 of the fish landing series

Mechanised

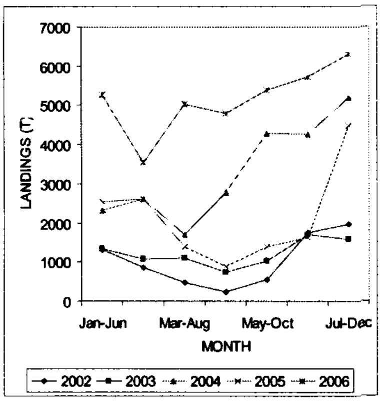


Figure 3.7(g)

Motorised

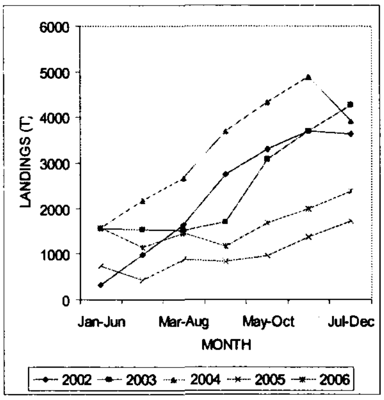


Figure 3.7(h)

Non-mechanised

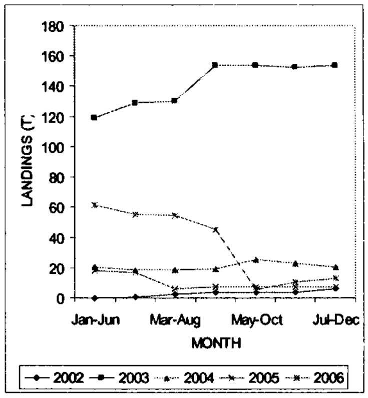


Figure 3.7(i)

Figure 3.8 Trend in the fishing boats operated at zone K6 during 2002-06

Mechanised

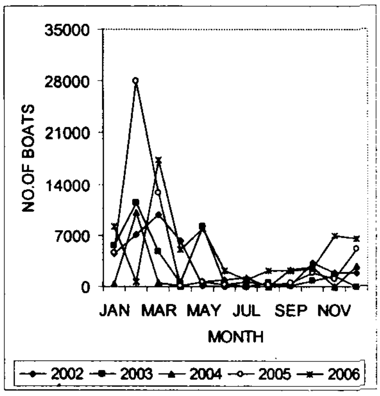


Figure 3.8(a)

Motorised

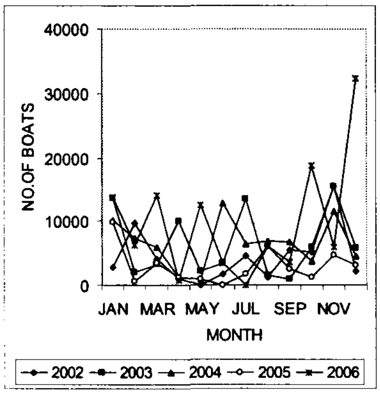


Figure 3.8(b)

Non-mechanised

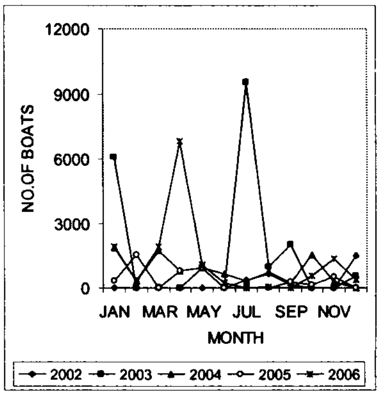


Figure 3.8(c)

Moving Average of order 3 of the fishing boats series

Mechanised	Motorised	Non-mechanised

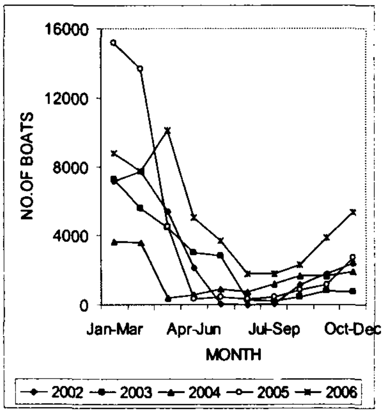


Figure 3.8(d)

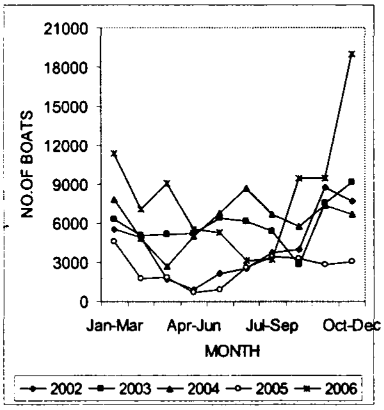


Figure 3.8(e)

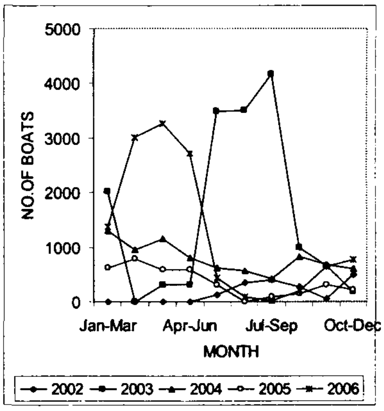


Figure 3.8(f)

Figure 3.9 Trend in the landing per boat at zone K6 during 2002-06

Mechanised	Motorised	Non-mechanised

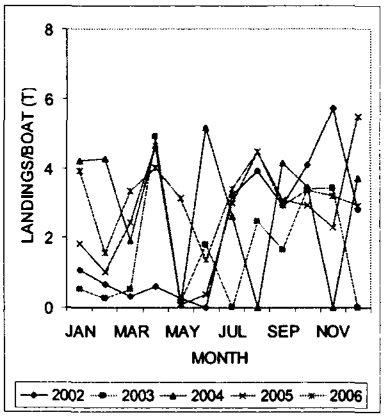


Figure 3.9(a)

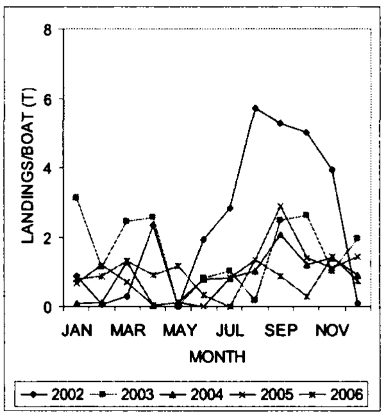


Figure 3.9(b)

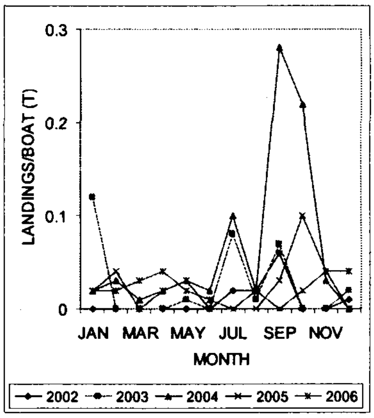


Figure 3.9(c)

Zone K8

In this zone nearly 50% of the landings is by motorised boats, 48% by mechanised boats and the remaining 2% is by non-mechanised boats. Since the zone contributes more than 20% of the state's total landing every year, it is one of the most important zones. Figure 3.10(a) to (i) indicate the behaviour of the total landings and the 3 and 6 term moving averages for the zone. The total landings are quite irregular in all the three sectors over the years. The moving average plots indicate that the landings by the mechanised and non-mechanised sector decrease over the months every year in contrary to the very slight increase in the motorised sector. Figure 3.11(a) to (f) exhibiting the behaviour of the number of boats operated indicate a totally irregular trend in all the three sector with an obvious dip towards the beginning of the monsoon season. Figure 3.12(a) to (c) for the landings per boat also exhibits high degree of variation.

Figure 3.10 Trend in the landings at zone K8 during 2002-06

Mechanised **Motorised** **Non-mechanised**

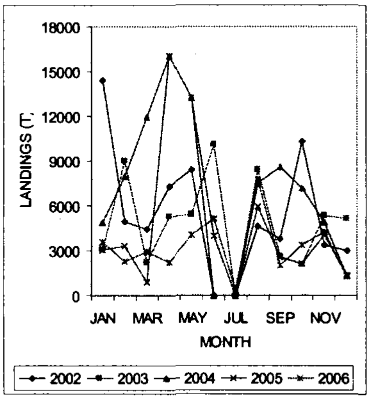


Figure 3.10(a)

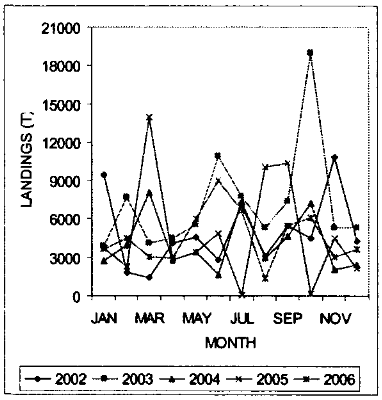


Figure 3.10(b)

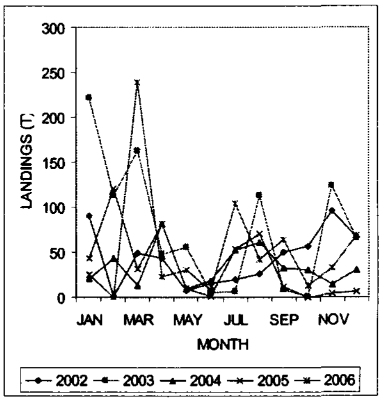


Figure 3.10(c)

Moving Average of order 3 of the fish landing series

Mechanised **Motorised** **Non-mechanised**

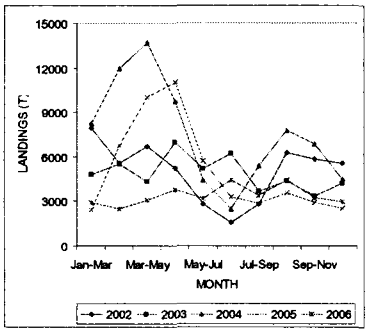


Figure 3.10(d)

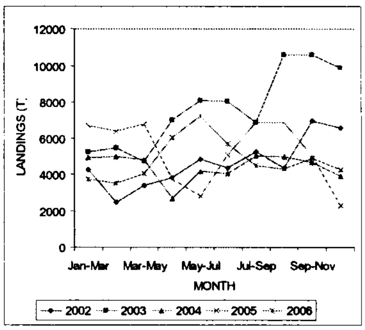


Figure 3.10(e)

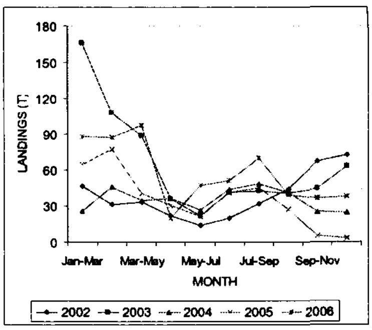


Figure 3.10(f)

Moving Average of order 6 of the fish landing series

Mechanised | **Motorised** | **Non-mechanised**

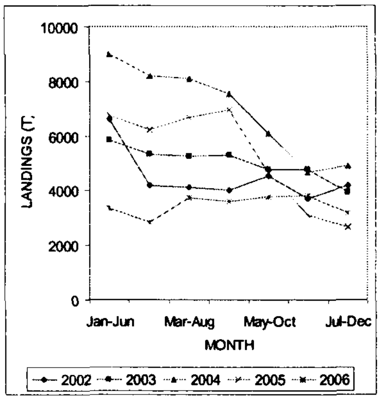


Figure 3.10(g)

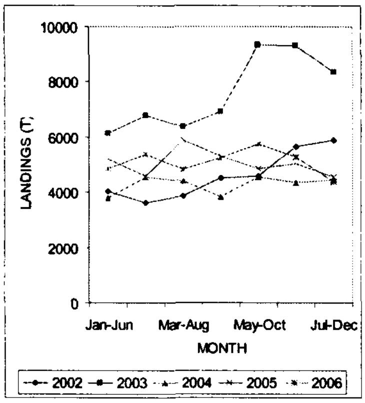


Figure 3.10(h)

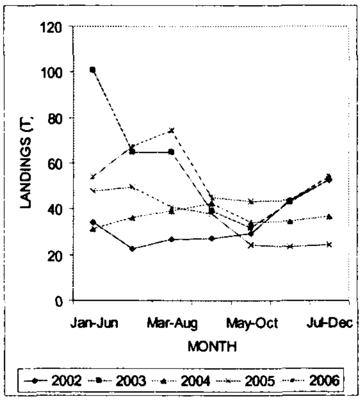


Figure 3.10(i)

Figure 3.11 Trend in the fishing boats operated at zone K8 during 2002-06

Mechanised | **Motorised** | **Non-mechanised**

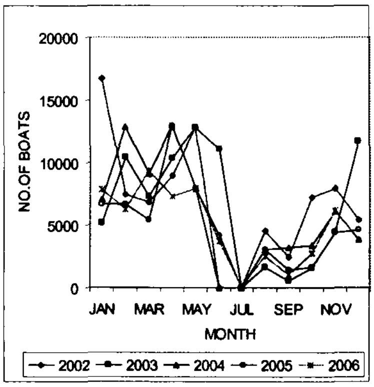


Figure 3.11(a)

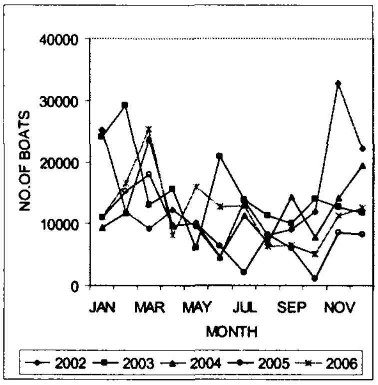


Figure 3.11(b)

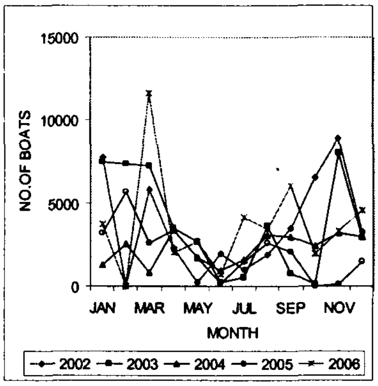


Figure 3.11(c)

Moving Average of order 3 of the fishing boats series

Mechanised | **Motorised** | **Non-mechanised**

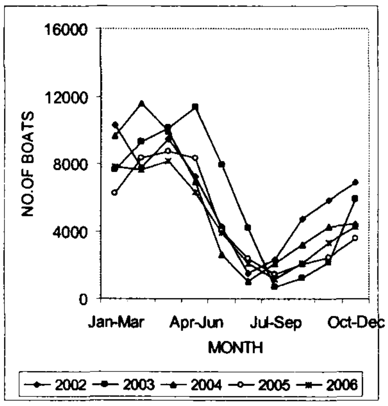


Figure 3.11(d)

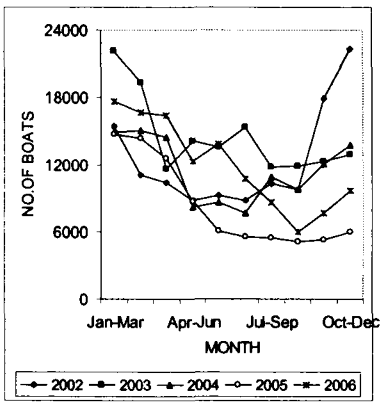


Figure 3.11(e)

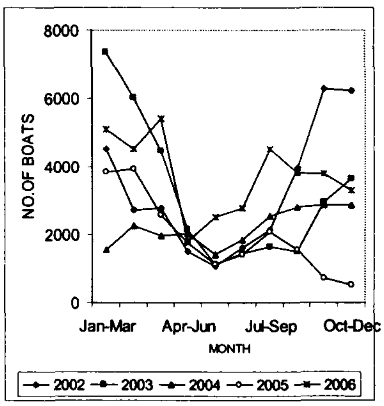


Figure 3.11(f)

Figure 3.12 Trend in the landing per boat at zone K8 during 2002-06

Mechanised **Motorised** **Non-mechanised**

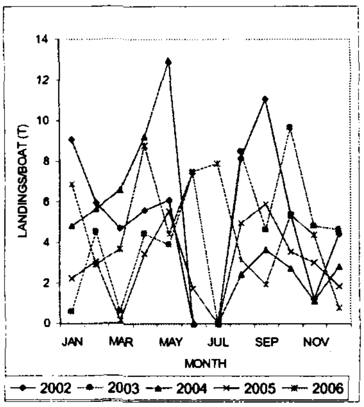


Figure 3.12(a)

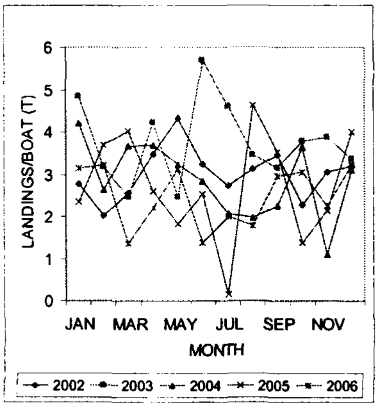


Figure 3.12(b)

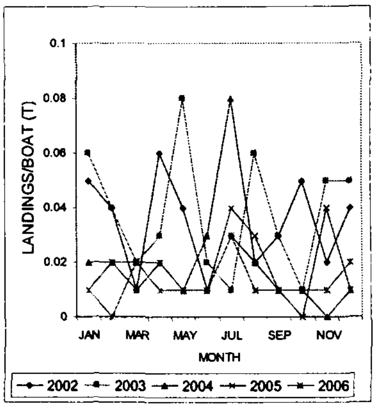


Figure 3.12(c)

Zone K11, Neendakara Fisheries Harbour

This is a single centre zone recording the highest landings, nearly 10% of the state's total over the period. Only motorised and mechanised boats operate at this centre. The landings are in general irregular. The 6 term moving averages, figure 3.13(e) and (f) indicate a slight increasing trend in the landings by mechanised boats while those of the motorised shows an increasing trend during 2002 and 03 and a reverse trend in 2004, 05 and 06. Figure 3.14(a) to (d) indicates that the number of boats of either kind exhibits a decreasing trend towards the beginning of the monsoon season. Figure 3.15(a) and (b) exhibit that the landings per boat also is quite irregular in this zone.

In all other zones for which graphs are not included, the behaviour of the landings and number of boats were more or less of the irregular forms described above.

Figure 3.13 Trend in the landings at K11 during 2002-06

Mechanised **Motorised**

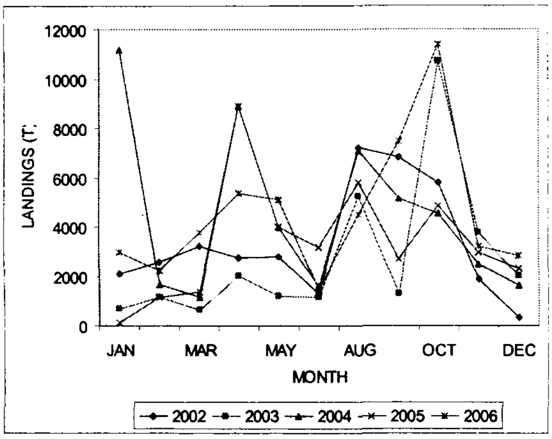


Figure 3.13(a)

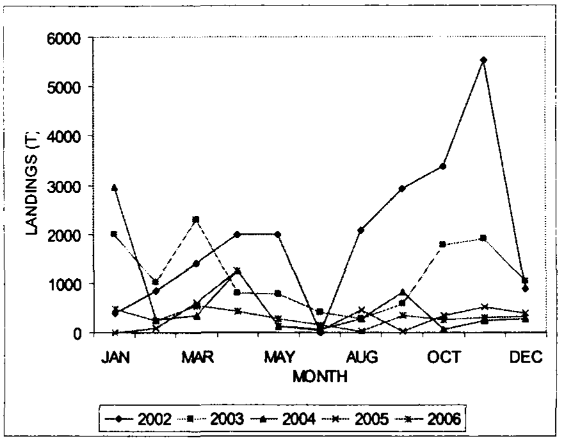


Figure 3.13(b)

Moving Average of order 3 of the fish landing series

Mechanised **Motorised**

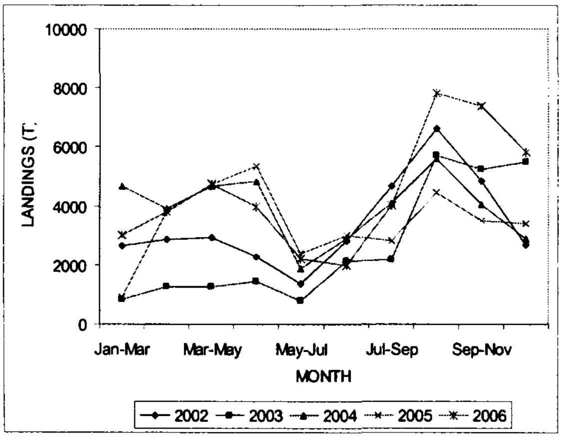


Figure 3.13(c)

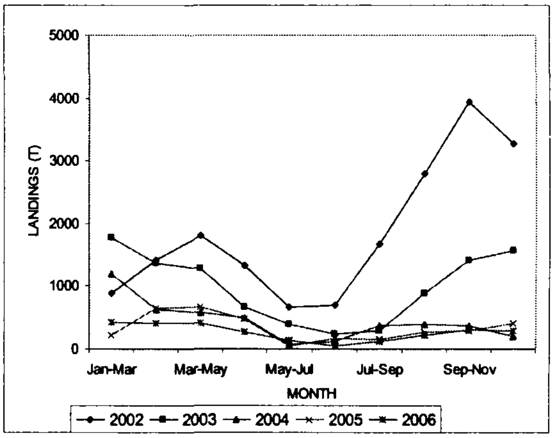


Figure 3.13(d)

Moving Average of order 6 of the fishing landing series

Mechanised **Motorised**

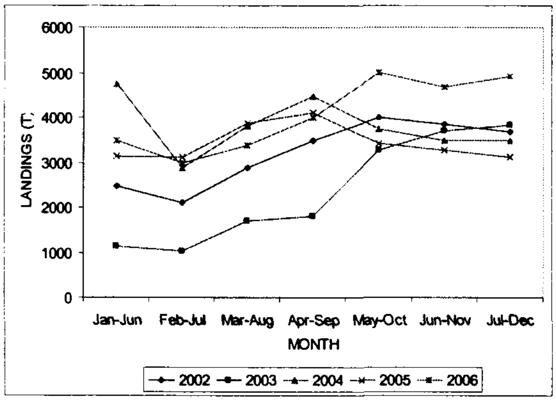


Figure 3.13(e)

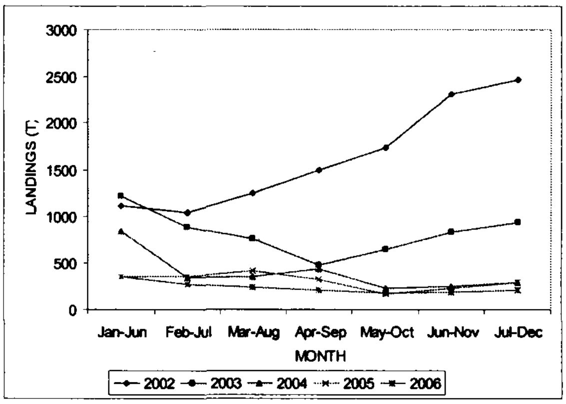


Figure 3.13(f)

Figure 3.14 Trend in the fishing boats operated at K11 during 2002-06

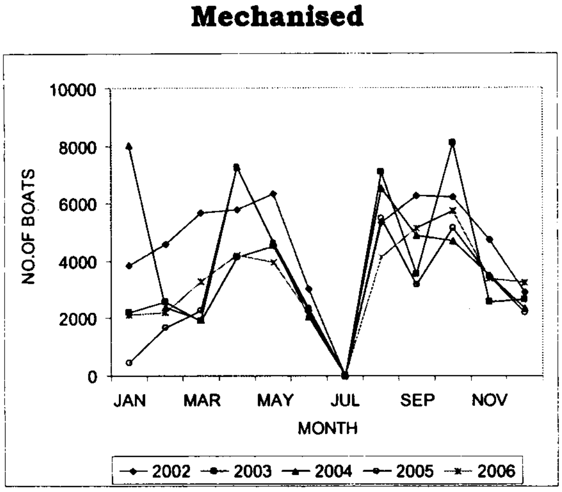


Figure 3.14(a)

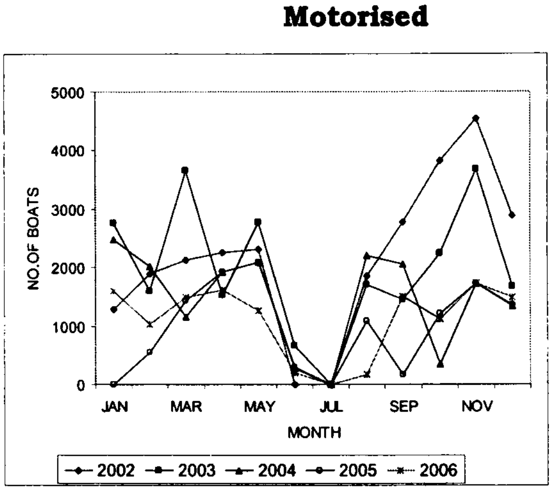


Figure 3.14(b)

Moving Average of order 3 of the fishing boats series

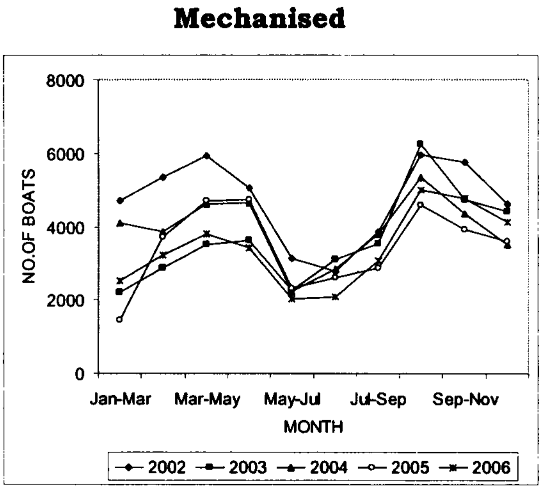


Figure 3.14(c)

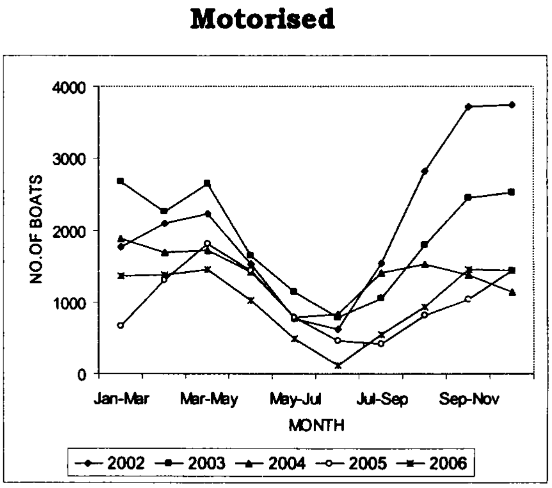


Figure 3.14(d)

Figure 3.15 Trend in the landing per boat at K11 during 2002-06

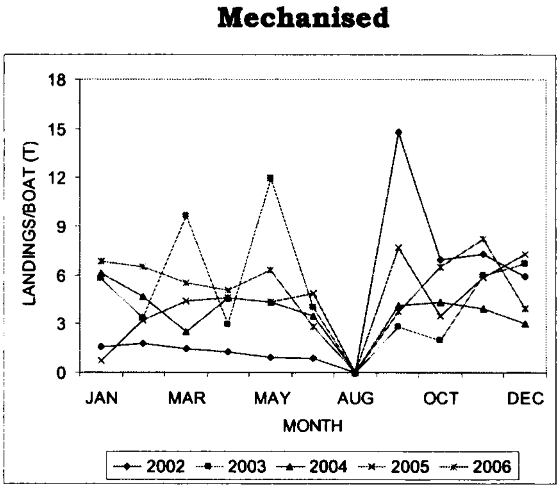


Figure 3.15(a)

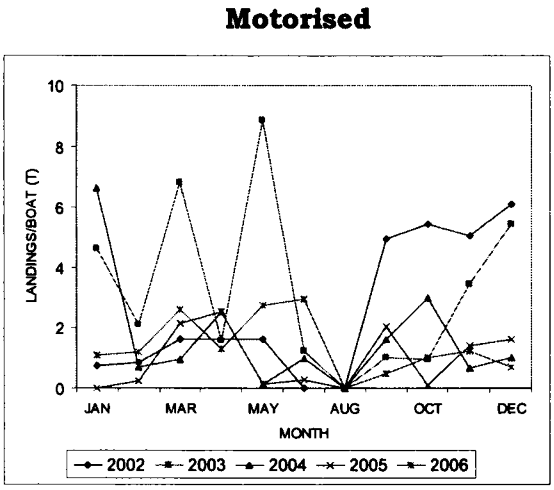


Figure 3.15(b)

3.5 Comparison of period wise landings at single centre zones

In this section we examine the variation if any between the landings in the two day-periods of observations. The total landings during the 6 hr period at the selected centre was estimated by multiplying the average landings per boat by the total number of fishing boats landed.

To test the difference between average quantity of fish landed at the forenoon and afternoon periods, 't' the conventional t-test is used.

Let y_{1i} and y_{2j} denote the landings and $\bar{y}_1$ and $\bar{y}_2$ the average landings during the afternoon (AN) and forenoon (FN) periods of observations respectively $(i = 1,2,...,n_1, j = 1,2,...,n_2)$. Then the test statistic under $H_o : \bar{y}_1 = \bar{y}_2$, the t-statistic is given by

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

which is assumed to follow a student t-distribution with $n_1 + n_2 - 2$ degrees of freedom under H_o.

where $s^2 = \frac{1}{n_1 + n_2 - 2}\left(\sum_{i=1}^{n_1}(y_{1i} - \bar{y}_1)^2 + \sum_{j=1}^{n_2}(y_{2i} - \bar{y}_2)^2\right)$ is the sample variance.

To apply the above test we require the estimate of variance for each period. Since meaningful estimates of variance are possible only if sufficient observations are available, the analysis is confined to single centre zones. Sample results of t-test together with the corresponding p-value based on the data at Neendakara and Sakthikulangara zones for the year 2006 are given in table 3.4. Figures 3.16 (a), (b) to 3.21 (a), (b) compare the number of boats landed and the total catch over the periods (AN and FN) for the single centre zones Neendakara, Sakthikulangara and Cochin fisheries harbour over the two years 2005 and 2006. The month for which data are not available are omitted in both the table and figures. Note that in table 3.4 the p-values for each month at Neendakara are greater than 0.10, in some cases as high as 0.90, indicating no significant difference between the landings over the two periods.

However, the p-values corresponding to Sakthikulangara are all less than 0.05 except three months, indicating that the landings during the two periods differ significantly except for the three months.

Figures 3.16(a) and (b) indicate that the number of boats landed at Neendakara during the two periods are more or less equal while the quantity of fish landed are explicitly different in the year 2005.

Table 3.4 Results of t-test for comparison of landings during afternoon and forenoon – 2006

Month	Neendakara				Sakthikulangara			
	Afternoon	Forenoon	t value	P value	Afternoon	Forenoon	t value	P value
Jan	17084	19799	-0.42	0.68	1108	10885	-7.26	0.00
Feb	16869	15727	0.21	0.84	6735	15102	-1.37	0.20
Mar	23624	16274	1.76	0.11	1277	11868	-9.91	0.00
Apr	24626	19157	0.84	0.42	3151	6571	-1.36	0.21
May	21622	20791	0.13	0.90	1576	6131	-3.50	0.01
Jun	7908	13682	-1.37	0.30	1677	16007	-4.71	0.01
Aug	22837	23226	-0.14	0.90	9931	22947	-2.09	0.07
Sep	30317	34939	-0.53	0.61	7030	14610	-2.82	0.01
Oct	26223	31386	-0.44	0.67	4080	19916	-4.58	0.00
Nov	13011	14516	-0.16	0.88	4681	15115	-6.01	0.00
Dec	15029	12478	0.54	0.60	3365	13004	-4.44	0.00

Figure 3.16 Comparison of boats landed and the average landings between afternoon and forenoon at Neendakara in 2005

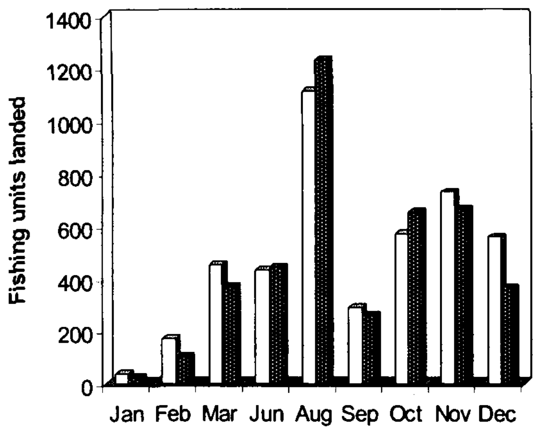


Figure 3.16(a)

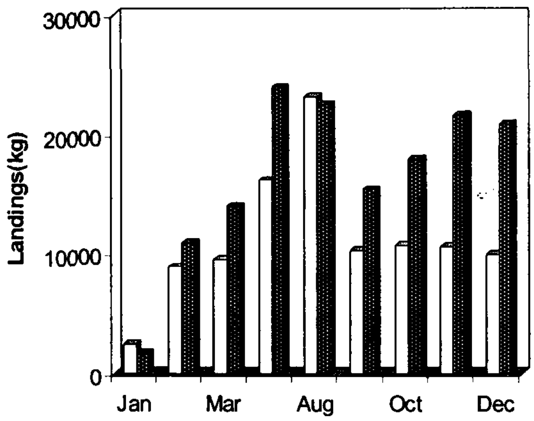


Figure 3.16(b)

Figure 3.17 Comparison of boats landed and the average landings between afternoon and forenoon at Neendakara in 2006

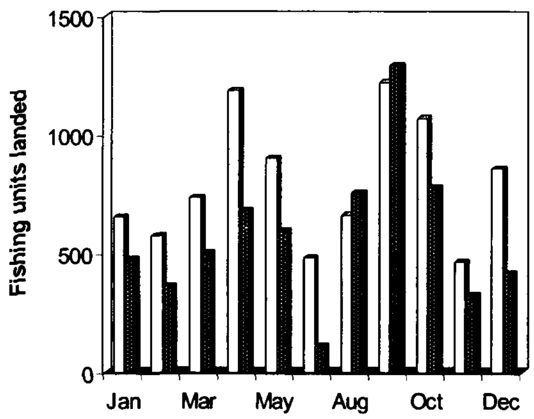


Figure 3.17(a)

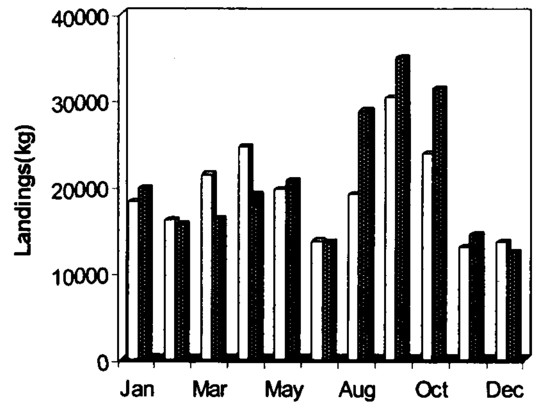


Figure 3.17(b)

Similarly figure 3.18(a) and 3.19(a) indicate that the wide disparity in the number of boats landed during the AN and FN periods. Note that except for the month of August, more than 90% of the boats land in the FN. Corresponding difference in the landings are also exhibited by the figure 3.18(b) and 3.19(b). The situation at Cochin fisheries harbour is also very similar to that at Sakthikulangara as exhibited by figure 3.20(a), (b) and 3.21(a), (b).

It is worth noting that the number of boats that land during the FN are relatively very high in most of the centres. However, the difference in the quantity of landings is not in proportion to the difference in the number of boats. This is due to the fact that the boats which land during the AN are those which have more capacity and spend more time for fishing and hence bring comparatively more fish.

Figure 3.18 Comparison of boats landed and the average landings between afternoon and forenoon at Sakthikulangara in 2005

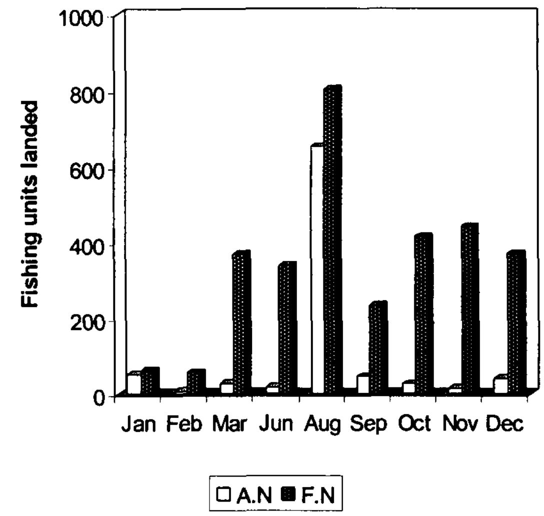


Figure 3.18(a)

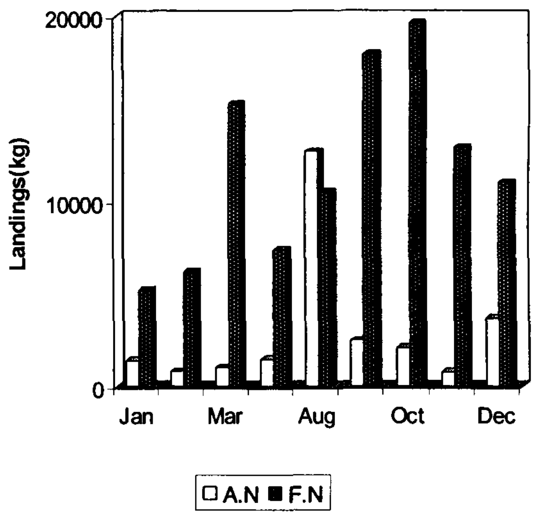


Figure 3.18(b)

Figure 3.19 Comparison of boats landed and the average landings between afternoon and forenoon at Sakthikulangara in 2006

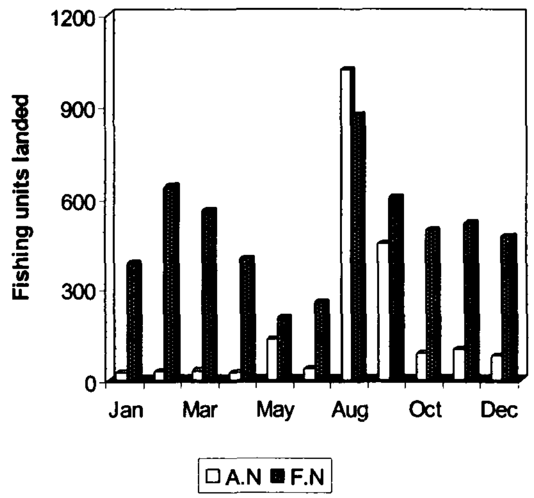


Figure 3.19(a)

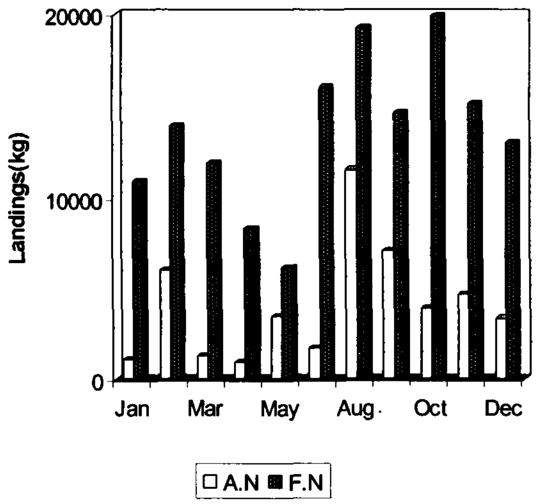


Figure 3.19(b)

Figure 3.20 Comparison of boats landed and the average landings between afternoon and forenoon at Cochin fisheries harbour in 2005

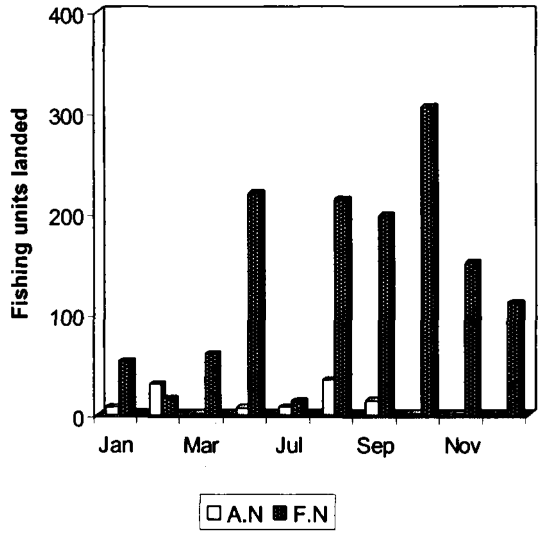


Figure 3.20(a)

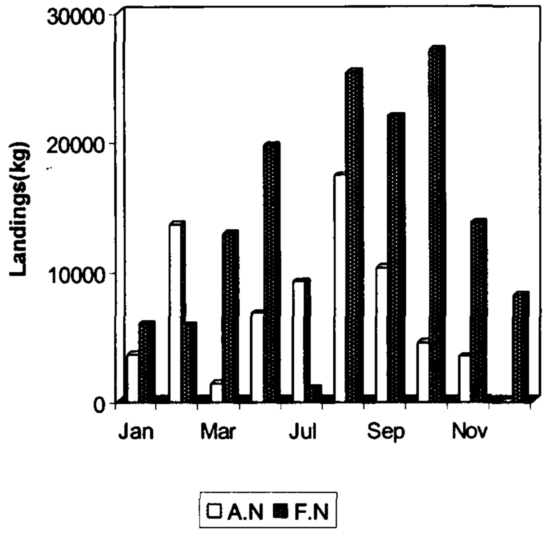


Figure 3.20(b)

Figure 3.21 Comparison of boats landed and the average landings between afternoon and forenoon at Cochin fisheries harbour in 2006

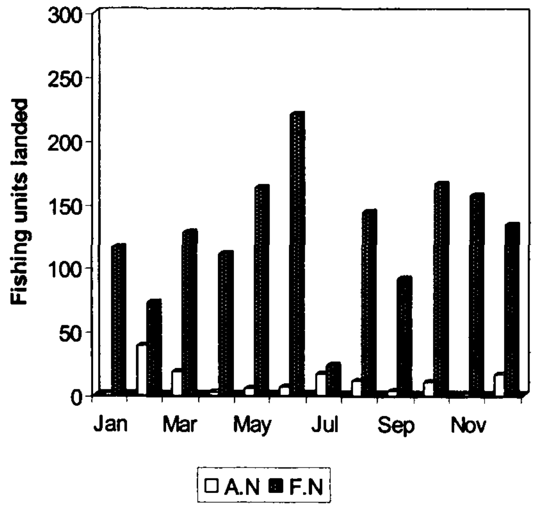


Figure 3.21(a)

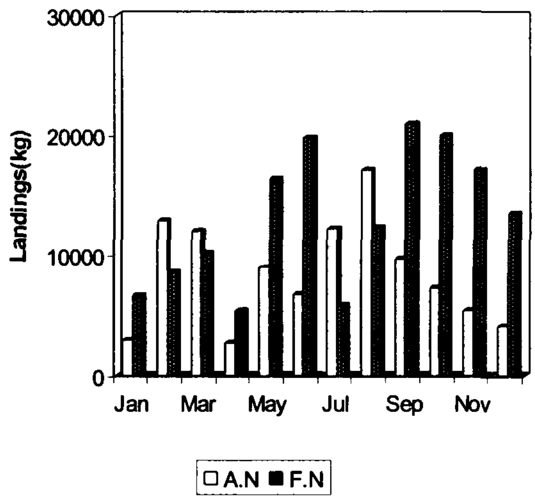


Figure 3.21(b)

From this analysis it can be concluded that to get a better estimate of the landings at a centre in a day (if fish landings takes place through out the day in the centre), it is necessary to collect the data both in the morning and afternoon hours.

3.5 Conclusion

The analysis performed in this chapter to assess the patterns and trends if any in the fishery data with regard to important basic characteristics of interest such as total landings, number of boats landed, the landings per boat and the period wise landings indicate that each of these characteristics are highly flexible and hence no predictable patterns or trends are possible. The flexibility may be attributed to several fishery dependent and independent factors. The fishery independent factors include meteorological and oceanographic variables, food availability etc, the fishery depended factors include craft used, nature, size and shape of the gears, their mesh size and the fishing effort, fishing grounds etc. In the existing fishery survey, information on some of the fishery dependent factors namely, craft used; the type of gear used, the fishing effort and fishing ground are available. Hence it may be concluded that;

(a) Fishery data is highly unstable and unpredictable.

(b) All sources of variations are to be carefully taken into account while using estimates for decision making.

(c) Estimation of variance may be given due consideration in analysing the data.

CHAPTER IV

Estimation of Variance Components by the Method of Nested Analysis

4.1 Introduction

An essential requirement of any survey is that a measure of precision be provided for each estimate derived from the survey data. It is well known that in almost all kinds of random sampling, the sample mean (total) is an unbiased estimator of the population mean (total). However, without a valid estimator of its variance, no reliable inferences can be made using it. Estimation of variance is a problem of main concern in most of the cases. The problem becomes more and more complex as the sampling scheme involves several stages and different sampling rules are adopted at different stages. Though the analysis of variance (ANOVA) technique is one of the standard procedures to be adopted to compute the variances, the special nature of the data sometimes does not yield to the traditional computation of the sum of squares. For complicated multistage data, special procedures are to be adopted.

The existing procedure of computing the variability in the estimate of the fish landings takes into account only the between day variation. As reflected in the analysis in chapter 3, the fish landings data is highly flexible and the variations may be due to each of the stages of the survey and the selection process adopted at each stage. The fishery survey data may be regarded as a survey involving as many as 3 stages. In stage 1, combining the landing centres and days of the month, a one dimensional structure is imagined and simple random sampling is applied to select the landing centre days. In stage 2, though all the gear types operated on a day are observed, when we take the month as a whole it is quite likely that a few of the gear types remain unobserved. Hence, assuming that T boat types operate in every month, T_i are observed on the i^{th} selected day, state 2 also can be regarded as involving a selection of T_i boat types out of T. The stage 3 is the selection of the boats which is done in a semi random manner. Thus the fish landings data collected for about 6 to 9 days in a month, in a zone is a three-stage completely mixed structure from which the sum of squares can not be extracted in the traditional way. In this chapter, we describe the nested model technique of constructing the ANOVA and estimating the variance components and apply the technique to estimate the components of variance due to each stage of sampling in the fish landings data. The general linear model for ANOVA of an unbalanced nested design and its analysis are described in section 4.2. Describing the nested structure of the marine fishery data, its analysis using a 3 stage model is described in section 4.3. Zone wise estimates of variance components as given by the method for the period 2004-06 are described in section 4.4.

4.2 Analysis of Hierarchical or Nested Data

Hierarchical or nested data arise very often in biological and chemical experiments as well as in multistage sampling. For example consider a biological experiment involving two or more factors each used at two or more levels. The experimental material will be such that there are as many rows as the level of the first factor each is then divided into as many columns as there are levels for the second factor, each cell is then sub divided into as many levels as there are levels for the third factor and so on. The resulting design will be a layout in the form of nesting. Similarly in the multistage sample surveys, the entire area under the survey would be divided into a large number of groups denoted by, say, z_1. Each of these z_1 groups, sub-divided into a large number of smaller groups, z_2 each of which in turn are split up into still smaller groups z_3 and so on until one arrives at the ultimate unit of sampling. The procedure that had been followed by first selecting a number of z_1 groups from the population and from each of these selected z_1 groups, selecting a number of z_2 groups, from each of the selected z_2 groups again a number of z_3 groups are chosen and so on until the ultimate units of sampling are reached, has been termed nested sampling by P.C. Mahalanobis (Ganguli, 1941). A nested design is said to be balanced if the number of subclasses at the h^{th} stage are the same for all the $(h-1)^{th}$ stage classes. In a multi stage sampling, balanced design refers to the case of the h^{th} stage units are of equal size. Otherwise the nested design is said to be unbalanced.

The hierarchical or nested data is usually analysed using a linear additive model. Three types of linear models viz.,

Linear fixed effect model,

Linear random effect model and

Linear mixed effect model

are available. A fixed effect model is one in which the effect of each of the factors are assumed to be fixed while the residual effect (error) alone is assumed to be a homoscedastic random variable. A random effect model is one in which except an overall general effect, the effects of all the factors are assumed to be independent random variables along with the residual effect (error). In a mixed effect model the effect of one or more of the factors are assumed to be fixed while that of the remaining factors and the residual effect (error) are assumed to be independent random variables. In nested designs random effect models are generally considered. The Analysis of variance (ANOVA) table based on the random effect model can be prepared in the same manner as in the case of fixed effect model. One of the chief distinction is that the mean sum of squares (MSS) in the ANOVA for a random effect model does not estimate the variance components directly. This is because of the mixing of the variance components due to nesting of one factor with in the other in several levels. It can be seen that each MSS turns out to be unbiased estimators of linear functions of the variance components. Hence to get separate estimators for the variance component corresponding to each factor we have to solve the linear functions of the variance components equated to the respective MSS. Though this method is theoretically very sound, in several practical cases it provides negative values as the estimate for the variance component causing difficulty in interpretation and decision making.

In a balanced hierarchical data, there is no much difficulty in the conduct of the tests of significance or in the estimation of variance components. Whereas, in the unbalanced case, the coefficients of a particular variance component in the expectation of MSS will vary from one mean square to another. However, there exists an orderly pattern among the coefficients of the variance components in the expected mean squares as given by equation (4.19), even in the unbalanced case. Ganguli (1941) was the first to give a detailed description of the algebraic calculation of expectations of MSS for the nested designs and estimation of variance components. Since the procedure is applied to the analysis of the fishery data for the first time, the brief details of the general procedure is also given below.

The general s-stage (level) unbalanced nested random effect model

Consider a s-stage nested design with I_h subclasses at level *h*, $h = 1,2,..,s+1$.

The model for an arbitrary s-level nested unbalanced design can be written as

$$Y_{i_1 i_2 i_3 \ldots i_{s-1} i_s i_{s+1}} = \mu + \alpha_1 + \beta_{2(1)} + \gamma_{3(2)} + \ldots + \varsigma_{s-1(s-2)} + \eta_{s(s-1)} + \varepsilon_{s+1(s)} \qquad \text{-----(4.1)}$$

where $Y_{i_1 i_2 i_3 \ldots i_{s-1} i_s i_{s+1}}$ is the observation corresponding to the i_{s+1}^{th} (ultimate) unit in the I_{s+1}^{th} class within the $I_s{}^{th}$ class within the $I_{s-1}{}^{th}$ class etc., $i_h = 1,2,\ldots,I_h$, $h = 1,2,\ldots,s+1$.

μ is the overall mean effect,

α_1 is a vector of I_1 components representing the effect of I_1 classes at level 1,

$\beta_{2(1)}$ is a vector of I_2 components representing the effect of I_2 classes at level 2,

$\gamma_{3(2)}$ is a vector of I_3 components representing the effect of I_3 classes at level 3,

.

.

.

$\xi_{s-1(s-2)}$ is a vector of I_{s-1} components representing the effect of I_{s-1} classes at level s-1,

$\eta_{s(s-1)}$ is a vector of I_s components representing the effect of I_s classes at level s and $\varepsilon_{s+1(s)}$ is a vector of I_{s+1} components representing the residual effect.

The following assumptions are made:

(i) μ is a fixed effect,

(ii) $\alpha_1, \beta_{2(1)}, \gamma_{3(2)}, \ldots, \eta_{s(s-1)}$ are the random effects, each normally distributed with mean 0 and variance $\sigma_1^2, \sigma_2^2, \ldots, \sigma_s^2$ respectively,

(iii) $\varepsilon_{s+1(s)}$ are independent identically distributed normal random variable with mean 0 and variance σ^2,

(iv) the effects of the various levels are independent, and homoscedastic.

For notational convenience let there be a super level denoted as 0-level to which all the 1st stage units belong and also assign the level $s+1$ for the class of ultimate observations, $I_0 = 1$ and $I_{s+1} = 1$.

N_h = the total number of sub classes at level h

n_h = the number of units in the sub class at level h.

$N_h = \sum_{i_1 i_2 \dots i_{h-1}} I_h$, $h = 1,2,\dots,s+1$, $N_0 = 1$ and $N_1 = I_1$

$n_{s+1} = 1$, $n_{h-1} = \sum_{i_h} n_h$, and $n_0 = N_{s+1}$ denote the total number of data values.

Let Y_h = the sum of the n_h observations in the subclasses at level h, Y_{s+1} being the individual observed values and Y_0 the grand total.

$$Y_{h-1} = \sum_{i_h} Y_h ,$$

$$\bar{Y}_h = \frac{Y_h}{n_h},$$

$h = 0,1,\dots,s+1$

S_0 is the correction factor and S_{s+1} is the sum of squares of all observations.

In the case of the random effect model (4.1), the hypotheses of interest to be tested are

$$H_1 : \sigma_1^2 = 0,$$

$$H_2 : \sigma_2^2 = 0,$$

$$\vdots$$

$$H_s : \sigma_s^2 = 0.$$

which can be written as linear hypotheses as in the fixed effect model as

H_1: all the I_1 components of α_1 are equal,

H_2: all the I_2 components of $\beta_{2(1)}$ are equal,

.

.

.

H_s : all the I_s components of $\eta_{s(s-1)}$ are equal.

Obviously, each H_h is a linear hypothesis of $(I_h - 1)$ degrees of freedom; $h = 1,2,...,s$. In the random effect model, since the α_1 , $\beta_{2(1)}$,..., are random variables, there is no relevance in estimating them.

Mean sum of squares and their expected values

From the model (4.1) as per the assumptions (i) to (iv), we get

$$V(Y_{i_1 i_2 \ldots i_s i_{s+1}}) = V(\alpha_1) + V(\beta_{2(1)}) + \ldots + V(\eta_{s(s-1)}) + V(\varepsilon_{s+1(s)})$$

If we denote $V(Y_{i_1 i_2 \ldots i_s i_{s+1}}) = \sigma^2$, we get

$$V(Y_{s+1}) = \sigma^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_{s+1}^2 \qquad \text{-----(4.2)}$$

Let

$$\bar{\theta}_{j,h} = \left(\frac{1}{n_h}\right) \sum_{i_{h+1},\ldots,i_s,i_{s+1}} \theta_{j(j-1)} \, , \; j = 1,2,\ldots,s+1 \, ; h = 0,1,\ldots,s$$

where θ denotes $\alpha, \beta, \gamma, \ldots, \varepsilon$ and

$$\bar{\theta}_{j,j} = \theta_{j(j-1)} ; \forall j$$

$$\bar{Y}_0 = \frac{1}{n_0} \sum_{i_1,\ldots,i_{s+1}} \left(\mu + \alpha_1 + \beta_{2(1)} + \ldots + \theta_{j(j-1)} + \ldots + \eta_{s(s-1)} + \varepsilon_{s+1(s)}\right)$$

$$= \mu + \sum_{i_1} \frac{n_1 \alpha_1}{n_0} + \sum_{i_1,i_2} \frac{n_2 \beta_{2(1)}}{n_0} + \ldots + \sum_{i_1,i_2,\ldots,i_h} \frac{n_h \theta_{h(h-1)}}{n_0} + \ldots + \sum_{i_1,\ldots,i_{s+1}} \frac{\varepsilon_{s+1(s)}}{n_0}$$

$$\bar{Y}_0 = \mu + \bar{\alpha}_{1,0} + \bar{\beta}_{2,0} + \ldots + \bar{\theta}_{h,0} + \ldots + \bar{\varepsilon}_{s+1,0} \qquad \text{-----(4.3)}$$

$$V(\bar{Y}_0) = V(\bar{\alpha}_{1,0}) + V(\bar{\beta}_{2,0}) + \ldots + V(\bar{\theta}_{h,0}) + \ldots + V(\bar{\varepsilon}_{s+1,0})$$

$$V(\bar{Y}_0) = \sum_{i_1} \frac{n_1^2}{n_0^2}\sigma_1^2 + \sum_{i_1,i_2} \frac{n_1^2}{n_0^2}\sigma_2^2 + \ldots + \sum_{i_1,i_2,\ldots,i_h} \frac{n_h^2}{n_0^2}\sigma_h^2 + \ldots + \frac{\sigma_{s+1}^2}{n_0} \qquad \text{-----(4.4)}$$

$$\bar{Y}_1 = \frac{1}{n_1} \sum_{i_2,\ldots,i_{s+1}} \left(\mu + \alpha_1 + \beta_{2(1)} + \ldots. + \theta_{h(h-1)} + \ldots + \eta_{s(s-1)} + \varepsilon_{s+1(s)}\right)$$

$$= \frac{1}{n_1}\left[n_1\mu + n_1\alpha_1 + \sum_{i_2} n_2\beta_{2(1)} + \ldots + \sum_{i_2,\ldots,i_h} n_h\theta_{h(h-1)} + \ldots + \sum_{i_2,\ldots,i_{s+1}} \varepsilon_{s+1(s)}\right]$$

$$= \mu + \alpha_1 + \bar{\beta}_{2,1} + \ldots + \bar{\theta}_{h,1} + \ldots + \bar{\varepsilon}_{s+1,1} \qquad \text{-----(4.5)}$$

$$V(\bar{Y}_1) = \sigma_1^2 + \sum_{i_2} \frac{n_2^2}{n_1^2}\sigma_2^2 + \ldots + \sum_{i_2,\ldots,i_h} \frac{n_h^2}{n_1^2}\sigma_h^2 + \ldots + \frac{n_1\sigma_{s+1}^2}{n_1^2} \qquad \text{-----(4.6)}$$

In general,

$$\bar{Y}_h = \frac{Y_h}{n_h} = \frac{1}{n_h} \sum_{i_{h+1},\ldots,i_{s+1}} \left(\mu + \alpha_1 + \beta_{2(1)} + \ldots. + \theta_{h(h-1)} + \ldots + \eta_{s(s-1)} + \varepsilon_{s+1(s)}\right)$$

$$\bar{Y}_h = \frac{1}{n_h}\left[n_h\mu + n_h\alpha_1 + n_h\beta_{2(1)} + \ldots + \sum_{i_{h+1}} n_{h+1}\theta_{h(h-1)} + \ldots + \sum_{i_{h+1},\ldots,i_{s+1}} \varepsilon_{s+1(s)}\right]$$

$$= \mu + \alpha_1 + \ldots + \bar{\theta}_{h,h} + \ldots + \bar{\varepsilon}_{s+1,h} \qquad \text{-----(4.7)}$$

$$V(\bar{Y}_h) = \sigma_1^2 + \sigma_2^2 + \ldots. + \sigma_h^2 + \frac{1}{n_h^2}\sum_{i_{h+1}} n_{h+1}^2\sigma_{h+1}^2 + \ldots + \frac{n_h\sigma_{s+1}^2}{n_h^2} \qquad \text{-----(4.8)}$$

for $h = 1,2,\ldots,s$.

The sum of squares (SS) due to the factors at the successive levels can be expressed as follows.

Sum of squares due to level 1,

$$SS_1 = \sum_{i_1}(\overline{Y}_1 - \overline{Y}_0)^2 \qquad \text{-----(4.9)}$$

using (4.3) and (4.5)

$$SS_1 = \sum_{i_1}\left[(\alpha_1 - \overline{\alpha}_{1,0}) + (\overline{\beta}_{2,1} - \overline{\beta}_{2,0}) + \dots + (\overline{\varepsilon}_{s+1,1} - \overline{\varepsilon}_{s+1,0})\right]^2 \qquad \text{-----(4.10)}$$

similarly, SS due to level 2

$$SS_2 = \sum_{i_1,i_2}(\overline{Y}_2 - \overline{Y}_1)^2$$

$$= \sum_{i_1,i_2}\left[(\overline{\beta}_{2,2} - \overline{\beta}_{2,1}) + (\overline{\gamma}_{3,2} - \overline{\gamma}_{3,1}) \dots + (\overline{\varepsilon}_{s+1,2} - \overline{\varepsilon}_{s+1,1}\right]^2 \qquad \text{-----(4.11)}$$

and the SS due to level *s*,

$$SS_s = \sum_{i_1,i_2\dots,i_{s+1}}(\overline{Y}_{(s+1)} - \overline{Y}_s)^2 = \sum_{i_1,i_2\dots,i_{s+1}}\left[(\varepsilon_{s+1,s} - \overline{\varepsilon}_{s+1,s})\right]^2 \qquad \text{-----(4.12)}$$

The expected values of the SS (4.9) to (4.12) can be derived easily, noting that α_1, $\beta_{2(1)}$, $\dots,\varepsilon_{s+1(s)}$ are independent. Further under the assumptions of the model it can be verified that

$$Cov\left[(\overline{\theta}_{j,h} - \overline{\theta}_{j,h-1}),(\overline{\theta}_{j,h-1} - \overline{\theta}_{j,h-2})\right] = 0, \ \forall h \,\&\, j \qquad \text{-----(4.13)}$$

Hence, we get that each of these SS are independently distributed as central χ^2 except for suitable scalar multiple.

Let $V_h = SS_h / d_h$ denote the Mean sum of squares (MSS) where $d_h = N_h - N_{h-1}$ is the degrees of freedom for the SS due to the h^{th} level, $h=1,2,\ldots,s+1$.

Using the property (4.13), the expected values of the MSS are given as follows.

$$E(V_1) \quad = \frac{1}{N_1 - N_0} E\left(\sum_{i_1} \left[(\alpha_1 - \bar{\alpha}_{1,0}) + (\bar{\beta}_{2,1} - \bar{\beta}_{2,0}) + \ldots + (\bar{\varepsilon}_{s+1,1} - \bar{\varepsilon}_{s+1,0}) \right]^2 \right)$$

using (4.4) and (4.6)

$$= \frac{1}{d_1} \left[\left\{ n_0 - \frac{\sum_{i_1} n_1^2}{n_0} \right\} \sigma_1^2 + \left\{ \sum_{i_1} \frac{\sum_{i_2} n_2^2}{n_1} - \frac{\sum_{i_2} n_2^2}{n_0} \right\} \sigma_2^2 + \ldots + \left\{ \sum_{i_1} \frac{\sum_{i_s} n_s^2}{n_1} - \frac{\sum_{i_s} n_s^2}{n_0} \right\} \sigma_s^2 + \sigma_{s+1}^2 \right]$$

----(4.14)

similarly

$$E(V_2) = \frac{1}{d_2} E\left(\sum_{i_1, i_2} \left[(\bar{\beta}_{2,2} - \bar{\beta}_{2,1}) + (\bar{\gamma}_{3,2} - \bar{\gamma}_{3,1}) + \ldots + (\bar{\varepsilon}_{s+1,2} - \bar{\varepsilon}_{s+1,1}) \right]^2 \right)$$

$$= \frac{1}{d_2} \left[\left\{ n_0 - \sum_{i_1} \frac{\sum_{i_2} n_2^2}{n_1} \right\} \sigma_2^2 + \ldots + \left\{ \sum_{i_2} \frac{\sum_{i_s} n_s^2}{n_s} - \sum_{i_1} \frac{\sum_{i_s} n_s^2}{n_1} \right\} \sigma_s^2 + \sigma_{s+1}^2 \right] \qquad \text{----(4.15)}$$

.

.

.

$$E(V_{s+1}) = \frac{1}{d_{s+1}} E\left(\sum_{i_1, i_2, \ldots, i_{s+1}} (\bar{Y}_{(s+1)} - \bar{Y}_s)^2 \right)$$

$$= E\left(\sum_{i_1, i_2, \ldots, i_{s+1}} \frac{1}{d_{s+1}} \left[(\varepsilon_{s+1(s)} - \bar{\varepsilon}_{s+1(s)}) \right]^2 \right) \qquad \text{----(4.16)}$$

$$= \sigma_{s+1}^2$$

and $E(V_{s+1}) = \sigma_{s+1}^2$ -----(4.17)

In general, the expected value of the h^{th} level MSS will be unbiased for $\sigma^2_{h(h+1)..s}$.

ie., $\sigma^2_{h(h+1)..s} = E(V_h) = \sigma_{s+1}^2 + k_{h,s}\sigma_s^2 + k_{h,s-1}\sigma_{s-1}^2 + \ldots + k_{h,h}\sigma_h^2$,

$h = 1, 2, \ldots, s$ ----- (4.18)

where $k_{h,j}$'s are given by

$$k_{h,j} = \frac{1}{d_h} \left[\sum_{i_h} \frac{\sum_{i_j} n_j^2}{n_h} - \sum_{i_{h-1}} \frac{\sum_{i_j} n_j^2}{n_{h-1}} \right], \text{ for } (j \geq h) \text{ and 0 otherwise,} \qquad \text{-----(4.19)}$$

where the summations are made over the j^{th}, h^{th} and $(h-1)^{th}$ subclasses. Thus we get the SS due to the h^{th} level follows $\chi^2 \sigma^2_{h(h+1)\ldots s(s+1)}$ with degrees of freedom d_h. Also as per condition (4.13) the SS are independently distributed.

Making use of these notations, the source of variation, degrees of freedom and sum of squares for the s-stage nested classification may be expressed as in Table 4.1.

In the ANOVA for non-nested models usually each MSS corresponds to the variation due to only one of the factors of comparison and hence the tests are conducted by taking the ratio of the MSS due to the factors to the MSS due to the error. However, in the case of general (unbalanced) nested model with random effects, as (4.18) reveals that the expected value of the MSS due to the h^{th} level is a linear combination of the variations due to all the factors from level h onwards. Hence, there is no meaning in taking the F-ratio as the ratio of a MSS due to the h^{th} level to the MSS due to error. Instead to test the significance of variance due to the h^{th} level, the F-ratio is to be taken as $F_h = \frac{V_h}{V_{h+1}}$, which follows F distribution with (d_h, d_{h+1}) degrees of freedom. If the p-value of the test is greater than α, H_h is accepted, otherwise it is rejected at level α. It is worth noting that the above F-test does not compare the variances due to the h^{th} and $(h+1)^{th}$ level as may be expected. Though in the case of balanced nested designs, under the null hypothesis, the F-test as described above is exact, it is not so when the design is unbalanced. This is due to the fact that in the unbalanced case the coefficients of the variance components in the expected values of the MSS due to the h^{th} and $(h+1)^{th}$ level are different and hence any difference in the MSS cannot be completely attributed to the level h. When the test leads to acceptance of $H_h : \sigma_h^2 = 0$, it is fine, but if the test leads to rejection of the null hypothesis we have to estimate σ_h^2.

Table 4.1 Analysis of variance table for the s-level nested data

Level and Number of classes		SS due to level h SS_h	df $d_h = N_h - N_{h-1}$	MSS $V_h = \frac{SS_h}{d_h}$	F-ratio $F_h = \frac{V_h}{V_{h+1}}$	Expected value of V_h with variance components and their coefficients					
h	N_h					σ_{s+1}^2	σ_s^2	σ_{s-1}^2	$\cdots$	σ_2^2	σ_1^2
1	N_1	$S_1 - S_0$	$N_1 - N_0$	V_1	F_1	1	$k_{1,s}$	$k_{1,s-1}$	$\cdots$	$k_{1,2}$	$k_{1,1}$
2	N_2	$S_2 - S_1$	$N_2 - N_1$	V_2	F_2	1	$k_{2,s}$	$k_{2,s-1}$	$\cdots$	$k_{2,2}$	
3	N_3	$S_3 - S_2$	$N_3 - N_2$	V_3	F_3	1	$k_{3,s}$	$k_{3,s-1}$	$\cdots$		
		$\vdots$	$\vdots$			$\vdots$	$\vdots$	$\vdots$			
$s-1$	N_{s-1}	$S_{s-1} - S_{s-2}$	$N_{S-1} - N_{S-2}$	V_{s-1}	F_{s-1}	1	$k_{s-1,s}$	$k_{s-1,s-1}$			
s	N_s	$S_s - S_{s-2}$	$N_S - N_{S-1}$	V_s	F_s	1	$k_{s,s}$				
$s+1$	N_{s+1}	$S_{s+1} - S_s$	$N_{S+1} - N_S$	V_{s+1}	-	1					
	Total	$S_{s+1} - S_0$	$N_{S+1} - 1$	-	-						

Note that (4.18) contains *s* equations in (*s+1*) unknowns, $\sigma_1^2,...,\sigma_{s+1}^2$. (4.17) gives the value of $\hat{\sigma}_{s+1}{}^2$ explicitly. Hence, solving the system of equations (4.18) using (4.17), we can find the estimates of all the variance components.

The estimates are given by

$$\hat{\sigma}_{s+1}{}^2 = V_{s+1}$$

and $\hat{\sigma}^2{}_h = \left(V_h - \sum_{j>h} k_{h,j}\hat{\sigma}_j{}^2\right) / k_{h,h}$, for $h = 1,2,...,s$ -----(4.20)

Confidence Interval for the mean

In the nested random effect model also an unbiased estimator of the (fixed effect) over all mean is given by $\bar{Y}_0$. The variance of this estimator in the general model is given by (4.4). Using the individual estimators of the variance components as given by (4.20) the variance of the mean $\sigma_{\bar{Y}}{}^2$ can be obtained. The 100(1-α)% confidence interval, $\bar{Y} \pm \sigma_{\bar{Y}} t_{\alpha/2}$ can be determined using the table value of the student t distribution for $n_0 - 1$ degrees of freedom.

Negative components of variance estimates

In a random effect nested model we have seen that the variance components can not be estimated directly. This is because of the mixing of the variance components due to nesting of one factor with in the other at successive levels. Equation (4.18) indicates that each mean sum of squares turns out to be unbiased estimators of linear functions of the variance components. Hence to get separate estimators for the variance component corresponding to each factor we have to solve the linear functions of the variance components equated to the respective mean sum of squares as given by (4.20). Though this method is theoretically very sound, in several practical cases it provides negative values as the estimate for the variance components causing difficulty in interpretation and decision making. The frequent occurrence of negative estimates for one or more of the variance components had limited the usefulness of the nested ANOVA. The problem of negative estimates of variance components has received a lot of attention in the literature (Nelder, 1954; Thompson, 1962; Anderson, 1965, Federer, 1968). It occurs due to the fact that the variability at a higher level is less than the variability at a smaller level, resulting in a negative estimate of variance (Fletcher and Underwood, 2002).

There have been several efforts to cope with the phenomenon of negative estimates of variance components. Ganguli (1941) suggest to conside the negative estimate of variance as zero. McHugh and Mieke (1968) attributed two possible reasons for negative estimates of variance.

(i) The model is incorrect.

(ii) The statistical noise obscuring the underlying situations.

The model incorrectness is attributed to the violation of one or more of the assumptions. Normally it occurs due to the factors being correlated. Anscombe (1948) and Nelder (1954) have done valuable work which adopts the incorrect model. McHugh and Mieke (1968) examined the possibility of incorrect model, by considering the assumptions of sampling from infinite populations are incorrect. The correctness of the model can be ascertained by re-examining the background of the data generation process. Once the model correctness is ascertained, the negative estimates would be due to statistical noise alone. The problem due to statistical noise can be got rid of by dropping factors one by one from the model.

The usual procedure of dealing with a negative estimator is to set it to zero and remove the corresponding factor from the model and recalculate the estimates for the remaining factors. If more than one estimate is negative, remove first the smallest. However, setting one of the variance components to zero is to alter the remaining estimates considerably in some cases rendering them biased.

Herbach (1959) used the maximum likelihood principle to obtain variance component estimators which are non-negative. The maximum likelihood estimates are not generally unbiased but they often have smaller variance than the unbiased estimators. Thompson (1962) used the restricted maximum likelihood (REML) estimation procedure for balanced data which yields non-negative estimates, by assuming the model to be correct. The REML estimators are defined in such a way that they never become negative.

Thompson and Moore (1963) described the following procedure called 'pool the minimum violator' to resolve the problem of negative estimates of variance components. Draw a tree graph with its vertices representing effects due to the various factors in the ANOVA. The graph is such that the root of the tree denotes the error sum of squares, and all the single factor effects at level 1 at the top. All the joint effects at the successive levels are arranged from top to bottom, with multiple effects at the same level indicated at the same heights. The sum of squares in the ANOVA are supposed to be in the same top-down order of their magnitude as shown by the graph. The sum of squares at the same height being non comparable. Arrange the mean sum of squares in the ANOVA in the same order of this tree graph. If none of them violates the order then all variance components are to have nonnegative estimates. If any one of them violates the natural order then the estimate of the corresponding variance component will become negative. The pooling method is to drop the corresponding factor from the model and combine the sum of squares due to this factor to the sum of squares just below this in the graph. The pooled mean sum of squares is the weighted average of the two mean sum of squares each weighted with its degrees of freedom. In the case of several mean sum of squares violating the natural order, pooling is to begin with the minimum among them. The process may be continued till all the mean sum squares become in the natural order.

In the following sections, we apply the technique of nested design to estimate the variance components due to the different stages in fish landings data.

4.3 Nested Analysis of Fishery Data

The figure 4.1 is a schematic of nested structure of marine fish landings data in India. The country is divided into different states. Each state is divided into several zones. Only one of the state is considered in the figure 4.1. Since the states and zones are considered independently and identically, nesting is considered from landing centre days onwards only.

Zone is taken as 0-level,

Landing centre day within zone – level 1. (Each zone consists of *ND* landing centre days out of which n are selected)

Gear type – level 2, T_i gears operated on the i^{th} day.

Boats of gear type j – level 3. M_{ij} boats of type j landed on the i^{th} day out of which m_{ij} boats observed.

Species landed – level 4 (ultimate units).

The three level nested model is

$$Y_{i_1 i_2 i_3 i_4} = \mu + \alpha_1 + \beta_{2(1)} + \gamma_{3(2)} + \varepsilon_{4(3)}; \qquad \text{----(4.19)}$$

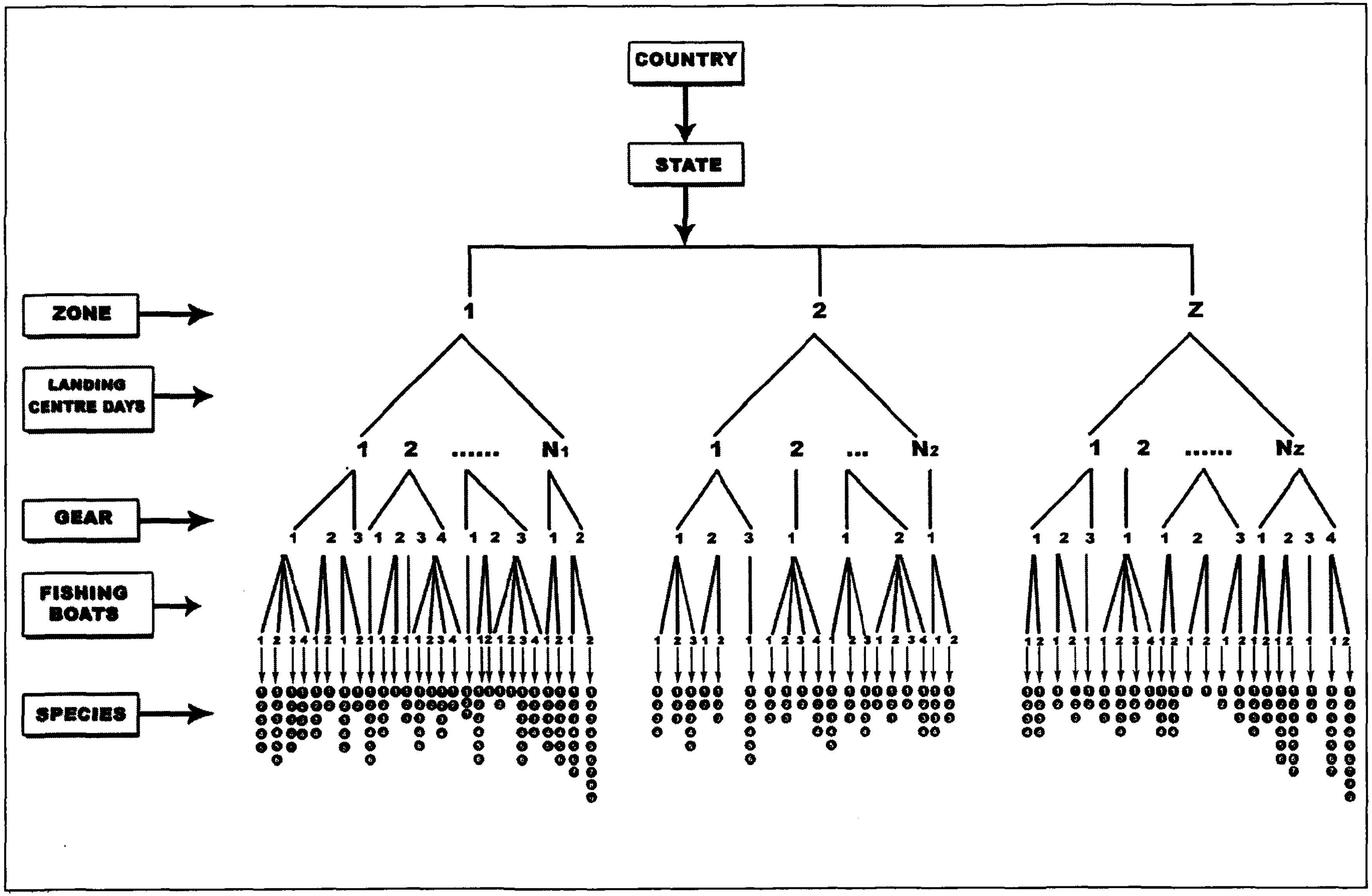
$$i_1 = 1,2...,n;\ i_2 = 1,2...,T_{i_1};\ i_3 = 1,2...,M_{i_1 i_2};\ i_4 = 1,2...,n_{i_1 i_2 i_3}.$$

where $Y_{i_1 i_2 i_3 i_4}$ is the quantity of fish of the i_4th species landed by the i_3^{th} fishing boat belonging to the i_2th fishing gear on the i_1th landing centre day, $n_{i_1 i_2 i_3}$ denotes the number of species observed in the i_3^{th} boat of the i_2th gear type on i_1th landing centre day.

Comparing with the general model (4.1) we have, $I_1 = n$, $I_2 = T_{i_1}$, $I_3 = M_{i_1 i_2}$ and $I_4 = n_{i_1 i_2 i_3}$

μ is the overall mean effect

Figure 4.1 Nested structure of marine fish landings data



α_1 is a vector of n components representing the landing centre day effect,

$\beta_{2(1)}$ is a vector of I_2 components representing the gear effect,

$\gamma_{3(2)}$ is a vector of I_3 components representing the boat effect,

and $\varepsilon_{4(3)}$ is a vector of I_4 components representing the residual effect.

The species and their quantity caught varies with respect to landing centre days, type of gear used, fishing units used, duration of fishing hours, depth in which gear was operated, time of fishing operation, climatic conditions etc. Hence, all the terms in the model (except μ) are assumed to be random variables. All the parameters are subjected to the assumptions under model (4.1).

The analysis of the fish landings data using the nested model (4.19) described in the above section of this chapter is made for each zone for the three year period 2004-06. The analysis was done using the 'nested procedure' and 'variance component procedure' in the Statistical Analysis System package. The 'nested procedure' performs a random-effects analysis of variance which is appropriate for a multistage nested sampling design. Rather than estimating the true variances due to each level by eliminating all the negative estimates, our objective here is to get an idea about the percentage contribution of each level to the total variance in each zone for each month. Due to paucity of space only the minimum representative outputs necessary for establishing our objective are included. Table (4.2) is the ANOVA table corresponding to the zone K1 for the year 2006. The last two columns of this table indicate the estimate of the variance components due to each level and its percentage share to the total variance. As indicated in section 4.3, the cases of all negative estimates of variances are taken as zero. Note that the estimate of variance due to day is zero for all the months and that due to boats is zero except for three months. In the three non-zero cases the percentage share were 17% in March, 11% in May and only 1% in July. The major shares of variation were due to gears and residual. The gear variation is as high as 67% in the month of March and the residual variation is as high as 69% in the month of September. The table indicates that the variation in the fish landings data is mainly due to gear-wise and residual variations while the contribution due to day-wise and boat-wise variations are negligible. Since, Table (4.2) is a typical representative of all other zones for the three years of analysis, the other ANOVA tables are not included. Instead tables indicating the average landings / species / gear / boat / day and its standard error and the percentage variation due to each levels as given by the ANOVA table are given in Tables (4.3(1)) to (4.3(15)) month-wise for each zone for the year 2006. The details of the months in which more than 10% contribution of variation are observed due to days and boats for each zone as given by the respective ANOVA tables for the 3 years are made and presented in Table (4.4). This table reveals that out of the 15 x 12 x 3 = 540 cases of study only about 13% cases exhibited more than 10% variance contribution due to days and boats. The zones K2, K5, K6 and K7 indicated more than 25% contribution due to day wise variation, while zones K6 and K12 indicated more than 25% contribution due to boat-wise variation. Tables 4.3(1) to 4.3(15) indicate that there are only a very few cases in which the contribution of the variation due to gears is less than 10%. The total number of such cases among the study cases is found to be less than 7%.

Table 4.2 Analysis of variance for the marine fish landings data for the zone K1 during 2006

Month	Source of variation	D.f.	Sum of squares	Mean Sum of squares	Variance components	
					Estimate	Percentage
January	Day	4	463.2	115.8	-12.5	0
	Gear	20	10301.0	515.0	47.0	51
	Boat	101	3058.9	30.3	-6.1	0
	Residual	168	7465.8	44.4	44.4	49
	Total	293	21289.0	72.7	91.5	100
February	Day	3	2615.6	871.9	-48.0	0
	Gear	14	29126.0	2080.4	188.2	63
	Boat	71	8656.3	121.9	9.5	3
	Residual	126	12538.0	99.5	99.5	33
	Total	214	52936.0	247.4	297.2	100
March	Day	5	40026.0	8005.3	-443.0	0
	Gear	16	317894.0	19868.0	1735.1	67
	Boat	104	134821.0	1296.4	443.9	17
	Residual	131	54607.0	416.8	416.8	16
	Total	256	547348.0	2138.1	2595.8	100
April	Day	3	1559.5	519.8	-288.8	0
	Gear	11	141491.0	12863.0	950.6	38
	Boat	66	53837.0	815.7	-311.2	0
	Residual	119	185591.0	1559.6	1559.6	62
	Total	199	382479.0	1922.0	2510.1	100
May	Day	30	4923.7	1641.2	-7.5	0
	Gear	9	12913.0	1434.8	117.7	35
	Boat	61	15116.0	247.8	37.0	11
	Residual	72	12746.0	177.0	177.0	53
	Total	145	45699.0	315.2	331.7	100
June	Day	30	1043.7	347.9	-90.6	0
	Gear	6	17347.0	2891.1	174.4	35
	Boat	37	6966.7	188.3	-39.1	0
	Residual	112	35550.0	317.4	317.4	65
	Total	158	60907.0	385.5	491.8	100

(contd.)

Month	Source of variation	D.f.	Sum of squares	Mean Sum of squares	Variance components	
					Estimate	Percentage
July	Day	4	9174.7	2293.7	-349.1	0
	Gear	17	215789.0	12693.0	1157.7	65
	Boat	90	60039.0	667.1	26.1	1
	Residual	146	88755.0	607.9	607.9	34
	Total	257	373758.0	1454.3	1791.7	100
August	Day	30	31885.0	10628.0	-3478.4	0
	Gear	15	2370642.0	158043.0	12534.0	39
	Boat	63	684944.0	10872.0	-3051.1	0
	Residual	160	3143872.0	19649.0	19649.0	61
	Total	241	6231343.0	25856.0	32183.0	100
September	Day	4	296266.0	74067.0	-1124.0	0
	Gear	13	1633701.0	125669.0	10691.0	31
	Boat	59	602399.0	10210.0	-5550.0	0
	Residual	118	2857441.0	24216.0	24216.0	69
	Total	194	5389807.0	27783.0	34906.0	100
October	Day	3	5999.6	1999.9	-23.8	0
	Gear	12	27794.0	2316.2	174.0	36
	Boat	75	12657.0	168.8	-60.7	0
	Residual	131	40963.0	312.7	312.7	64
	Total	221	87415.0	395.5	486.7	100
November	Day	3	2299.6	766.5	-45.1	0
	Gear	12	27072.0	2256.0	178.9	31
	Boat	63	14278.0	226.6	-68.7	0
	Residual	121	48338.0	399.5	399.5	69
	Total	199	91988.0	462.3	578.4	100
December	Day	3	4465.8	1488.6	-13.4	0
	Gear	10	17030.0	1703.0	113.0	34
	Boat	69	10870.0	157.5	-25.2	0
	Residual	128	28154.0	220.0	220.0	66
	Total	210	60520.0	288.2	333.0	100

Table 4.3 Zonal monthly estimate of average landings (in kg.) and the standard error for the year 2006

Zone K1

Month	$\hat{\bar{Y}}$	S.E($\hat{\bar{Y}}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	8.0	0.6	0	51	0	49
Feb	12.2	1.7	0	63	3	33
Mar	24.6	4.8	0	67	17	16
Apr	22.9	0.0	0	38	0	62
May	17.5	3.4	0	35	11	53
Jun	14.8	0.0	0	35	0	65
Jul	23.0	0.0	0	65	1	34
Aug	64.0	0.0	0	39	0	61
Sep	60.5	24.3	0	31	0	69
Oct	16.6	2.9	0	36	0	64
Nov	14.3	2.2	0	31	0	69
Dec	14.3	2.8	0	34	0	66

4.3 (1)

Zone K2

Month	$\hat{\bar{Y}}$	S.E($\hat{\bar{Y}}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	10.3	1.4	0	38	9	53
Feb	13.4	3.7	2	28	0	70
Mar	45.5	27.3	3	14	0	83
Apr	30.9	41.5	42	2	0	56
May	25.4	11.2	29	20	13	38
Jun	26.7	3.5	0	36	0	54
Jul	264.0	165.6	61	0	0	39
Aug	273.5	68.9	0	51	0	49
Sep	153.9	87.1	6	45	5	44
Oct	31.2	9.4	14	0	0	86
Nov	16.8	2.0	1	6	2	91
Dec	24.4	0.0	0	59	0	41

4.3 (2)

Zone K3

Month	$\hat{Y}$	S.E($\hat{Y}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	14.5	6.7	0	35	0	65
Feb	10.6	3.1	0	63	0	37
Mar	16.4	5.1	8	21	0	71
Apr	42.0	12.1	3	8	0	89
May	30.4	13.5	17	1	0	82
Jun	17.4	5.0	0	31	0	69
Jul	28.6	10.9	0	44	0	56
Aug	31.5	39.2	54	0	10	36
Sep	29.8	10.3	4	16	0	80
Oct	23.1	16.8	6	30	0	64
Nov	8.6	7.6	53	1	0	46
Dec	18.2	15.1	10	11	0	79

4.3 (3)

Zone K4

Month	$\hat{Y}$	S.E($\hat{Y}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	59.7	0.0	0	79	14	7
Feb	133.7	54.1	0	80	0	20
Mar	168.0	233.2	18	66	11	5
Apr	52.7	88.5	0	99	1	0
May	26.9	23.7	0	60	37	3
Jun	201.6	0.0	0	89	1	10
Jul	172.6	60.7	0	73	0	27
Aug	332.4	80.8	1	26	0	73
Sep	267.4	31.4	0	78	0	22
Oct	74.9	0.0	0	81	5	14
Nov	46.8	0.0	0	89	1	10
Dec	81.7	35.6	0	91	1	8

4.3 (4)

Zone K5

Month	$\hat{\bar{Y}}$	S.E($\hat{\bar{Y}}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	3.5	1	0	15	0	85
Feb	4.9	5.4	0	76	8	16
Mar	5.6	2.9	8	0	0	92
Apr	3.3	0.0	0	0	7	93
May	4.6	0.0	0	0	0	100
Jun	30.3	133.8	0	100	0	0
Jul	123.9	291.0	84	2	14	0
Aug	2.0	0.4	8	0	0	92
Sep	2.4	0.4	0	22	0	78
Oct	6.3	1.3	0	0	31	69
Nov	125.6	1700.6	98	2	0	0
Dec	3.8	1.1	8	2	0	90

4.3 (5)

Zone K6

Month	$\hat{\bar{Y}}$	S.E($\hat{\bar{Y}}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	294.4	528.7	72	15	7	6
Feb	309.4	279.7	0	75	20	5
Mar	101.6	84.6	0	78	11	11
Apr	310.3	233.0	0	87	0	13
May	124.4	107.9	0	87	8	5
Jun	113.6	106.6	28	5	0	67
Jul	2412.6	1396.0	60	11	29	0
Aug	1396.2	528.4	0	69	4	27
Sep	1482.7	483.5	0	72	9	19
Oct	572.0	549.6	83	0	8	9
Nov	214.0	448.3	0	94	4	2
Dec	87.0	145.8	0	94	1	5

4.3 (6)

Zone K7

Month	$\hat{Y}$	S.E($\hat{Y}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	145.2	757.7	0	98	2	0
Feb	66.8	15.8	0	7	0	93
Mar	109.6	0.0	0	88	11	1
Apr	283.5	346.7	34	0	0	66
May	78.1	73.4	17	56	0	27
Jun	401.6	745.9	43	34	15	8
Jul	22.1	0.0	0	25	0	75
Aug	2.4	0.5	15	0	0	85
Sep	140.6	235.0	89	0	11	0
Oct	120.8	287.8	0	98	2	0
Nov	71.6	164.7	88	0	0	12
Dec	58.6	198.2	62	0	7	31

4.3 (7)

Zone K8

Month	$\hat{Y}$	S.E($\hat{Y}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	111.9	57.3	0	82	2	16
Feb	109.0	85.9	0	77	16	7
Mar	64.5	9.0	0	46	5	49
Apr	90.5	42.9	0	59	0	41
May	112.0	36.8	0	75	7	18
Jun	149.6	44.8	0	46	1	53
Jul	90.9	0.0	0	92	5	3
Aug	130.5	65.9	0	32	0	68
Sep	203.3	73.6	0	64	5	31
Oct	295.5	95.6	0	39	9	52
Nov	76.7	35.2	0	54	4	42
Dec	72.1	0.0	0	67	1	32

4.3 (8)

Zone K9

Month	$\hat{Y}$	S.E($\hat{Y}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	161.7	94.7	0	75	1	24
Feb	127.6	104.2	0	98	1	1
Mar	80.7	30.1	0	62	19	19
Apr	116.4	172.6	0	73	4	23
May	283.8	0.0	0	97	1	2
Jun	84.1	29.6	0	17	0	83
Jul	227.7	180.5	0	94	3	3
Aug	9.7	3.8	0	25	0	75
Sep	346.5	176.6	0	73	0	27
Oct	280.1	0.0	0	94	3	3
Nov	59.3	0.0	0	87	1	12
Dec	34.1	6.4	0	5	0	95

4.3(9)

Zone K10

Month	$\hat{Y}$	S.E($\hat{Y}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	96.5	292.6	0	98	0	2
Feb	43.7	0.0	0	95	0	5
Mar	11.9	3.0	7	0	0	93
Apr	14.1	6.5	0	67	0	33
May	85.2	0.0	0	75	3	22
Jun	27.4	2.8	0	0	0	100
Jul	55.8	38.4	29	0	0	71
Aug	12.3	97.6	100	0	0	0
Sep	62.3	13.8	0	0	0	100
Oct	295.5	95.6	0	39	9	52
Nov	17.5	5.4	9	0	0	91
Dec	36.0	32.2	0	88	1	11

4.3 (10)

Neendakara

Month	$\hat{Y}$	S.E($\hat{Y}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	115.3	29.1	0	87	1	12
Feb	117.5	34.9	0	71	8	21
Mar	114.9	0.0	0	71	2	27
Apr	141	25.0	0	68	1	31
May	9	38.1	0	65	7	28
Jun	138.5	50.6	0	92	2	6
Jul	99.7	38.4	29	0	0	71
Aug	279.0	88.6	0	39	11	50
Sep	170.1	0.0	0	70	0	30
Oct	150.3	46.2	0	74	0	26
Nov	83.5	30.0	0	55	0	45
Dec	85.6	14.3	0	70	3	27

4.3 (11)

Sakthikulangara

Month	$\hat{Y}$	S.E($\hat{Y}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	95.4	16.6	0	54	9	37
Feb	162.4	40.2	0	60	8	32
Mar	95.8	13.5	0	38	40	22
Apr	65.7	9.2	0	57	9	34
May	50.8	13.9	6	8	4	82
Jun	140.5	39.6	0	38	32	30
Jul	55.8	38.4	29	0	0	71
Aug	197.7	22.4	0	20	0	80
Sep	108.3	30.9	0	16	34	50
Oct	168.7	29.8	0	30	13	57
Nov	116.9	21.6	0	24	23	53
Dec	102.3	30.9	0	64	5	31

4.3 (12)

Cochin

Month	$\hat{Y}$	S.E($\hat{Y}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	67.4	10.4	0	72	7	21
Feb	149.1	84.1	0	94	3	3
Mar	215.0	57.3	0	87	2	11
Apr	70.9	0.0	0	69	0	31
May	138.6	23.8	0	35	0	65
Jun	183.1	24.1	0	80	0	20
Jul	712.8	660.5	0	92	4	4
Aug	213.9	0.0	0	73	0	27
Sep	172.0	57.0	0	59	2	39
Oct	164.0	17.0	0	80	5	15
Nov	150.4	23.3	0	67	11	22
Dec	99.9	0.0	0	73	3	24

4.3 (13)

Munambam

Month	$\hat{Y}$	S.E($\hat{Y}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	140.9	20.8	0	34	0	66
Feb	153.6	281.6	0	1	9	20
Mar	131.4	24.7	0	31	0	69
Apr	224.1	50.5	0	34	0	66
May	263.0	21.5	0	60	0	40
Jun	280.7	65.8	0	41	0	59
Jul	634.4	415.1	0	67	25	8
Aug	550.0	249.3	0	89	0	11
Sep	385.3	1277.7	0	92	0	8
Oct	256.7	204.9	0	33	0	67
Nov	90.5	51.8	0	16	0	84
Dec	109.8	37.8	0	11	0	89

4.3 (14)

Vypin

Month	$\hat{Y}$	S.E($\hat{Y}$)	Percentage of total variance			
			Day	Gear	Boat	Residual
Jan	108.7	38.0	0	58	4	32
Feb	181.9	48.4	0	80	3	17
Mar	171.7	56.5	0	24	23	53
Apr	167.7	43.0	0	18	0	82
May	273.0	104.1	0	18	0	82
Jun	318.7	87.0	0	37	0	63
Jul	1268.6	0.0	0	83	9	8
Aug	447.2	475.5	78	0	6	16
Sep	509.6	118.2	0	14	0	86
Oct	331.8	0.0	0	43	9	48
Nov	150.0	25.6	0	58	7	35
Dec	70.0	22.7	0	25	0	75

4.3(15)

Table 4.4 Zone-Month exhibiting more than 10% variation due to days and boats

Zone	Day variation >= 10%			Boat variation >= 10%		
	2004	2005	2006	2004	2005	2006
K1	-	-	-	Apr	Mar	Mar, May
K2	Jan, Jul, Oct	Jul, Nov	Apr, May, Jul, Oct	-	Mar	May
K3	Mar, May, Jun, Sep	Jan, Feb, Mar, Nov	May, Aug, Nov, Dec	-	-	Aug
K4	May, Oct	Jan	Mar	-	-	Jan, Mar, May
K5	Apr, May, Jun, Aug	Jan, Feb, Jul, Sep	Jul, Nov	Jan	Oct, Dec	Jul, Oct
K6	Apr, May, Sep, Oct	Mar, Dec	Jan, Jun, Jul, Oct	Mar, May, Oct	Jan, Mar, Sep, Oct	Feb, Mar, Jul

Contd.

Zone	Day variation >= 10%			Boat variation >= 10%		
	2004	2005	2006	2004	2005	2006
K7	Jan, Jul, Nov	Jul, Nov	Apr, May, Jun, Aug, Sep, Nov, Dec	Sep, Nov	Jun, Nov	Mar, Jun, Sep
K8	Jul, Dec	Sep	-	Apr, Dec	Jan, Mar	Feb
K9	-	-	-	Jan	-	Mar
K10	Jan, Nov	-	Jul, Aug	-	Aug	-
K11	-	-	-	-	Jan	Aug
K12	-	-	-	Feb, Mar, Apr, Dec	Jan, Feb, Mar, Sep, Oct, Nov	Mar, Jun, Sep. Oct, Nov
K13	-	Mar	-	Feb, Sep, Dec	Jan, Feb, Sep, Oct	Nov
K14	-	-	-	-	Jul	Jul
K15	Feb	Mar, Sep	Aug	Mar, Sep	Jul, Sep	Mar

4.4 Conclusion

In the present study, the estimate and its variance of average landings / species/ gear/ boat / day for each zone-month stratum was found out using the three stage nested random effect model. The estimate of variance provided by the nested model in the present study accounts the variability at each stages of the design. Further, the proportion of total variance accounted at each level in the fish landings data was computed. The marine fisheries resources are dynamic and subject to fluctuations due to fishery dependant as well as fishery independent factors. The fishery independent factors include meteorological and oceanographic variables, food availability etc, the fishery depended factors include craft used, nature, size and shape of the gears, their mesh size and the fishing effort, fishing grounds etc. Similarly, the fish landings also varies with respect to the above factors. The between boats variation was due to the difference in the fishing effort, fishing capacity of the boat, fishing ground in which the boat was operated etc.

The reason for very low gear wise variation may be due to the operation of very few gears that too of the same type during the selected days. It was observed that there exist a variety of craft-gear boats operating in each zone, resulting in wide variation between the landings. The between gear variation was due to the different type of gears operated in the selected day. The between days variation was mainly due to the presence or absence of one or the other types of crafts landed. Since there exist a lot of variation between the type of gear used, it is necessary to sample more number of each type of fishing units on the selected day. The precision of the estimate of fish landings can be improved by incorporating the stratification by type of gear, either in the selection stage of the units or in the estimation stage.

The nested analysis made in this chapter reveals that the total variance of the fish landing data is the sum of variation due to each of the levels – between days, between gears, between boats and the residual. The magnitude of the contribution of each of the level may differ depending on several factors such as the seasonality, climatic conditions, fishing practices etc. Hence, the component-wise contributions of variance must be taken into account in analysing the data in any zone in any month. Further, it is observed that the major contribution to the variance is always due to gear types. Hence, any analysis of the data should essentially take into account the variation due to gear types.

However, for simplicity of analysis ignoring the between days and between boats variation may not cause very high discrepancy in most of the cases. In cases of the between day variation comparatively high, it is necessary either to observe more number of days. In the case of between boats variation are comparatively high, the boats may be stratified according to the fishing effort and capacity.

CHAPTER V

Estimation of Fish Landings - Post-Stratified Designs

5.1 Introduction

In the concluding section of chapter 2, we have noted that the low sampling fraction and the method of estimation of variance, ignoring all sources of variation other than the variation due to days are the two major limitations of the existing design. With an objective to rectify these limitations, in this chapter, we introduce two modified sampling designs based on post stratification of the data – one applicable to the single centre zones and the other for multi centre zones.

In the case of single centre zones, the sampling fraction can be increased to any desired level by simply increasing the number of days of observing the centre. In the case of multi-centre zones, the population has a two dimensional structure with landing centres in a zone in one dimension and days in a month in the other dimension. Hence an increase in the sampling fraction can be achieved very easily in a very scientific way, if this two dimensional structure of the population is taken into account in the sampling design. In the proposed design for multi-centre zone, this two dimensional structure is maintained by selecting the landing centres and observing them for a few selected number of days.

The nested analysis in chapter 4, reveals that the major share of the variation in the landings is due to gear types. Hence, any sampling design to be efficient must essentially be capable of adequately accounting for the variation due to gear types. This is possible only if the landings are stratified according to gear types. Due to the lack of knowledge of complete frame for the fishing boats of different gear types, we cannot adopt this technique in advance. So, in the proposed designs, post-stratification technique is adopted.

The first new design developed in section 5.2 is the same as the two stage sampling currently followed by CMFRI with the only modification that the data is post-stratified according to the observed gear types. The second new design developed in section 5.3 is a three stage design which retains the two dimensional structure of the population over space and time and adopts post-stratification based on gear types at the third stage.

5.2 Post-Stratified Design for Single Centre Zones

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. Since gear wise variation is very significant, this may result in an over or under estimation amounting to very high discrepancy in the estimate. In the absence of advance information on the gear types operated at the centre, post-stratification is the only way to get more reliable estimates. Post-stratification involves assignment of units into different category after selection of the sample.

Mehrotra (1993) gave a scheme for post-stratification in two stage sampling on the basis of the sampled second stage units. He demonstrated it empirically using a simulated data on area under high yielding varieties of wheat crop in a holding as the character under study. The PSUs being the number of villages and the SSUs the cultivator's holdings growing high yielding variety of wheat in a district. He suggested that this scheme not only provides estimates of the character under study according to the strata, but also improves the precision of the estimate pooled over the strata compared to the conventional non-stratified procedure. In the present study, we used this scheme to estimate the gear wise landings in a month in a zone.

The New Design I

This design is intended for single centre zones. Out of the N landing centre days n are selected adopting SRSWOR. Specified number of boats are selected on each selected day and landings are observed. The boats landed and those selected for observation on each day are stratified according to gear types. One important modification of the new design is regarding the 24 hr duration of the day as a single unit instead of regarding it as divided into three periods. The existing procedure of selecting the boats for observation as described in section 2.2.3 may be substituted with the following new strategy. The count of the boats landed to be recorded continuously through out the day. For recording the catch details, boats are to be selected at intervals of 15 to 20 minutes during the time of field visit with priority for distinct gear types if available. In the case of no new gear types available, priority is to be given to get at least two or three boats of the same gear. The night landings are to be considered only for recording count by enquiry in the forenoon of the following day of visit. Treating the 24 hour duration of a day into a single unit will give more freedom to the field staff to ensure adequate representation to each distinct gear type operated on the day. The forenoon session of the following day can be mainly targeted to select new gear types as well as to give adequate representation to the already noted gear types. At the end of the day, the number of boats landed and the catches recorded are post-stratified according to the gear types. It is true that, T, the number of distinct gear types will vary with respect to the days of observation. Hence, T number of post-strata is taken as the total number of distinct gear types over the observed days in the month. The resulting design can be regarded as a two stage random sampling with post-stratification at the second stage with day as PSU and boat of specific gear type as SSU. To analyse the data we follow, the existing procedure itself coupled with the procedure for post-stratification by Mehrotra (1993).

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 landed on i^{th} day.

Analogous to the existing estimator (2.5), an unbiased estimator of the total fish landings by the j^{th} type gear is given by

$$_{post}\hat{Y}_j \quad = \frac{N_{(j)}}{n_{(j)}} \sum_{i=1}^{n_{(j)}} \frac{M_{i(j)}}{m_{i(j)}} \sum_{k=1}^{m_{i(j)}} y_{ik(j)}, \qquad \text{-----(5.1)}$$

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 (5.1) $y_{ik(j)}$ is the species-wise sum of the observations on the day. There is no period-wise summation as in (2.5) since the day is treated as a single unit.)

Estimate of the total landings for the zone in a month is given by

$$_{post}\hat{Y} \quad = \sum_{j=1}^{T} \frac{N_{(j)}}{n_{(j)}} \sum_{i=1}^{n_{(j)}} \frac{M_{i(j)}}{m_{i(j)}} \sum_{k=1}^{m_{i(j)}} y_{ik(j)} \qquad \text{-----(5.2)}$$

Mehrotra (1993) had shown that the estimator of the type given in (5.2) is unbiased for the population total and its variance $_{post}\hat{Y}$ is given by,

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

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

$$+N^2 \sum_{j=1}^{T} \frac{1}{nN}\left(1+\frac{1-w_{(j)}}{nw_{(j)}}\right) \sum_{i=1}^{N_{(j)}}\left(M_i^2\left(w'_{i(j)}\right)\left(\frac{1}{m_i}-\frac{1}{M_i}\right)+\frac{1-w'_{i(j)}}{m_i^2}\right) S_{i(j)}^2 \qquad (5.3)$$

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

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

$$S_{(j)}^2 = \frac{1}{N_{(j)} - 1} \sum_{i=1}^{N_{(j)}} \left[Y_{i(j)} - \bar{Y}_{(j)} \right]^2 \tag{5.4}$$

$$S_{(jj')} = \frac{1}{N_{(jj')} - 1} \sum_{i=1}^{N_{(jj')}} \left[Y_{i(j)} - \bar{Y}_{(j)} \right] \left[Y_{i(j')} - \bar{Y}_{(j')} \right] \tag{5.5}$$

$$S_{i(j)}^2 = \frac{1}{M_{i(j)} - 1} \sum_{k=1}^{M_{i(j)}} \left[Y_{ik(j)} - \bar{Y}_{i(j)} \right]^2 \tag{5.6}$$

Note that (5.3) 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

$$\begin{aligned} \hat{V}\left({}_{post}\hat{Y}\right) &= \sum_{j=1}^{T} N^2 \left(\frac{1}{n} - \frac{1}{N} \right) w_{(j)} s_{(j)}^2 \\ &+ \sum_{j \neq j'}^{T} N_{(j)} N_{(j')} \left(\frac{n_{(jj')}}{n_{(j)} n_{(j')}} - \frac{n_{jj'}^2}{n_{(j)} n_{(j')} N_{(jj')}} \right) s_{(jj')} \\ &+ \sum_{j=1}^{T} \frac{N_{(j)}}{n_{(j)}} \sum_{i=1}^{n_{(j)}} M_{i(j)}^2 \left(\frac{1}{m_{i(j)}} - \frac{1}{M_{i(j)}} \right) s_{i(j)}^2 , \end{aligned} \qquad \text{-----(5.7)}$$

where

$$s_{(j)}^2 = \frac{1}{n_{(j)} - 1} \sum_{i=1}^{n_{(j)}} \left(y_{i(j)} - \bar{y}_{(j)} \right)^2 \qquad \text{-----(5.8)}$$

$$s_{(jj')} = \frac{1}{n_{(jj')} - 1} \sum_{i=1}^{n_{(jj')}} \left(y_{i(j)} - \bar{y}_{(j)}\right)\left(y_{i(j')} - \bar{y}_{(j')}\right) \qquad \text{-----(5.9)}$$

$$s_{i(j)}^2 = \frac{1}{m_{i(j)} - 1} \sum_{k=1}^{m_{i(j)}} \left(y_{ik(j)} - \bar{y}_{i(j)}\right)^2 \qquad \text{-----(5.10)}$$

are the unbiased estimators of the respective population mean squares.

Equation (5.2) gives the estimator and (5.7) the estimate 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 is also well accounted. The procedure can also provide gear-wise estimators using equation (5.1).

An Empirical Illustration

The marine fish landings data at Cochin fisheries harbour during the year 2004 was used for the illustration. The fish landings data was collected by using the existing two stage design. 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 design I. In the new design, post stratification is 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 stratum. 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 have constructed 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 would have directly used for the estimation purpose. In this illustration, since, we focus on the gear-wise estimates, 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 is made for a month. Then, the number of each gear type landed is simulated by assuming that it follows a multinomial distribution. This information is used for the estimation. The estimates obtained by the existing design and the design I are given in Table 5.1 below.

Table 5.1 Comparison of existing design and design I

	Existing design			Design I		
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

* Data on the total number of crafts landed were not available for the months of May and June

Note that the CV based on the new design are less than that of the existing design for all months. In some months the reduction in the CV was even greater than 10%. Since the post-stratified estimator performed well in all months, this new design I and the corresponding estimators are recommended for use in single centre zones.

5.3 Post-stratified Design II for Multi-centre Zones

In this section we propose a more general design mainly intended for multi-centre zones divided into different strata. The proposed design is a three stage stratified procedure with stratification at the first stage and post-stratification at the third stage. Retaining the two-dimensional structure of the fish landings data, the new design is intended to ensure higher sampling fraction for each stratum and take into account the variability due to each source.

The New Design II

The proposed sampling design is a three stage stratified sampling design, where the stratification is done over space and time, with landing centres as PSUs, days of a month as SSUs and the fishing boats as third stage units (TSU). The space stratum is similar to the existing design for multi-centre zones, while the PSUs are landing centres instead of the existing landing centre days. The PSUs, SSUs and TSUs are regarded as independently selected according to SRSWOR as is done in the existing case. Here also, the 24 hr duration of a day is taken as a single unit for observation. The design is described in detail below.

Consider a multi-centre zone with N landing centres and D fishing days in a month. This can be regarded as a two dimensional population of ND units, N units along the first dimension (space) and D units along the second dimension (time). The N landing centres be divided into L non-overlapping strata each of size N_h $(h=1,2,...L)$ such that $\sum_{h=1}^{L} N_h = N$. From the h^{th} stratum, n_h landing centres are selected by SRSWOR such that $\sum_{h=1}^{L} n_h = n$. Also select d_h days out of the D days adopting SRSWOR. Each of the selected n_h centres are observed for d_h days in the h^{th} stratum. On each day, the boats arriving at the centre are observed at intervals of 15 to 20 minutes, treating the whole day as a single unit as in design 1 described in section 5.2. The night landings are accounted only for recording the count. At the end of the day, the data on count and catch are stratified according to the gear types. As in design I, let T denote the number of distinct gear types observed in a month, then the boats landed and landings recorded are post-stratified into T strata. Let m_{hijk} and M_{hijk} denote the number of boats of k^{th} gear type observed and landed on j^{th} day at the i^{th} landing centre in the h^{th} stratum.

Then let,

$$m_{hij} = \sum_{k=1}^{T} m_{hijk} \text{ and } M_{hij} = \sum_{k=1}^{T} M_{hijk}$$

$$m_{hi} = \sum_{j=1}^{d_h} m_{hij} \text{ and } M_{hi} = \sum_{j=1}^{D_h} M_{hij}$$

$$m_h = \sum_{i=1}^{n_h} m_{hi} \text{ and } M_h = \sum_{i=1}^{N_h} M_{hi}$$

$$m = \sum_{h=1}^{L} m_h \text{ and } M = \sum_{h=1}^{L} M_h$$

Let

Y_{hijkl} be the total quantity of fish landed by the l^{th} fishing unit of k^{th} gear type on the j^{th} day at the i^{th} centre of the h^{th} stratum (As in the previous design Y_{hijkl} is the species-wise sum of the observations on the day)

$Y_{hijk} = \sum_{l=1}^{M_{hijk}} Y_{hijkl}$

$Y_{hij} = \sum_{k=1}^{T} Y_{hijk}$

$Y_{hi} = \sum_{j=1}^{D_h} Y_{hij}$

$Y_h = \sum_{i=1}^{N_h} Y_{hi}$

The above notations in lower case letters denote the corresponding factors in the sample and notations with bar represents the corresponding means.

Analogous to the estimator (2.6), the estimator of the total fish landings for the zone in a month under the new design, $\hat{Y}_M$, is given by

$$\hat{Y}_M = \sum_{h=1}^{L} \frac{N_h D_h}{n_h d_h} \sum_{i=1}^{n_h} \sum_{j=1}^{d_h} \sum_{k=1}^{T} \frac{M_{hijk}}{m_{hijk}} \sum_{l=1}^{m_{hijk}} y_{hijkl} \qquad \text{-----(5.11)}$$

The gear estimator is given by

$$\hat{Y}_{MK} = \sum_{h=1}^{L} \frac{N_h D_h}{n_h d_h} \sum_{i=1}^{n_h} \sum_{j=1}^{d_h} \frac{M_{hijk}}{m_{hijk}} \sum_{l=1}^{m_{hijk}} y_{hijkl}$$

Then

$$E(\hat{Y}_M) = E_1 E_2 E_3 (\hat{Y}_M)$$

where E_1, E_2 and E_3 are expectations with respect to the successive stages of selection of landing centres, days and fishing boats respectively. The first two are assumed to be SRSWOR while the third is post-stratified. Then,

$$E(\hat{Y}_M) = E_1E_2E_3\left(\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{D_h}{d_h}\sum_{j=1}^{d_h}\sum_{k=1}^{T}\frac{M_{hijk}}{m_{hijk}}\sum_{l=1}^{m_{hijk}}y_{hijkl}\right)$$

$$= E_1E_2\left(\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{D_h}{d_h}\sum_{j=1}^{d_h}E_3\left[\sum_{k=1}^{T}\frac{M_{hijk}}{m_{hijk}}\sum_{l=1}^{m_{hijk}}y_{hijkl}\right]\right) \qquad \text{-----(5.12)}$$

Now

$$E_3\left(\sum_{k=1}^{T}\frac{M_{hijk}}{m_{hijk}}\sum_{l=1}^{m_{hijk}}y_{hijkl}\right) = E_3\left(M_{hij}\sum_{k=1}^{T}W_{hijk}\bar{y}_{hijk}\right)$$

$$= M_{hij}\sum_{k=1}^{T}W_{hijk}\bar{Y}_{hijk}$$

$$= M_{hij}\bar{Y}_{hij}$$

$$= Y_{hij} \qquad \text{-----(5.13)}$$

where $\bar{y}_{hijk}$ is the mean of the k^{th} post-stratum which is an unbiased estimator of $\bar{Y}_{hijk}$ and $W_{hijk} = \frac{M_{hijk}}{M_{hij}}$ is the k^{th} stratum weight. Using (5.13) in (5.12), we get

$$E(\hat{Y}_M) = E_1\left(\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{D_h}{d_h}\sum_{j=1}^{d_h}E_2(Y_{hij})\right)$$

$$E(\hat{Y}_M) = E_1\left(\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{D_h}{d_h}\sum_{j=1}^{d_h}\left(\frac{1}{D_h}\sum_{j=1}^{D_h}Y_{hij}\right)\right)$$

$$= \sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}E_1(Y_{hi})$$

$$= \sum_{h=1}^{L}\frac{N_h}{n_h}\left(\sum_{i=1}^{n_h}\frac{1}{N_h}\sum_{i=1}^{N_h}Y_{hi}\right)$$

$$= \sum_{h=1}^{L}N_h\bar{Y}_h$$

$$= Y \qquad \text{-----(5.14)}$$

Therefore $\hat{Y}_M$ is an unbiased estimator of Y, where Y is the total fish landings in a zone in a month.

Variance of the estimator

As per the sampling design, in a zone, there are three stages of sample selection with post-stratification in the third stage. Hence, the variance of the estimator has to take into account all the three stages of selection.

The variance of $\hat{Y}_M$ (Desraj, 1971) is given by

$$V(\hat{Y}_M) \qquad = V_1E_2E_3(\hat{Y}_M) + E_1V_2E_3(\hat{Y}_M) + E_1E_2V_3(\hat{Y}_M), \qquad \text{----(5.15)}$$

where the subscripts 1,2 and 3 indicate that the expectation and variance are taken with respect to the 1st, 2nd and 3rd stages of sampling respectively.

Now taking the first term of (5.15),

$$V_1E_2E_3(\hat{Y}_M) \quad = V_1E_2E_3\left(\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{D_h}{d_h}\sum_{j=1}^{d_h}\sum_{k=1}^{T}\frac{M_{hijk}}{m_{hijk}}\sum_{l=1}^{m_{hijk}}y_{hijkl}\right)$$

$$= V_1\left(\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{D_h}{d_h}\sum_{j=1}^{d_h}E_2\left(\sum_{k=1}^{T}\frac{M_{hijk}}{m_{hijk}}E_3\left(\sum_{l=1}^{m_{hijk}}y_{hijkl}\right)\right)\right) \qquad \text{by (5.13)}$$

$$= V_1\left(\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}D_hE_2\left(\sum_{j=1}^{d_h}Y_{hij}\right)\right)$$

$$= V_1\left(\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}Y_{hi}\right)$$

$$= \sum_{h=1}^{L}N_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)S_h^2, \qquad \text{----(5.16)}$$

where

$$S_h^2 \qquad = \frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(Y_{hi}-\bar{Y}_h\right)^2,$$

since the first stage units selection is according to SRSWOR.

Now the second term of (5.15) is

$$E_1V_2E_3(\hat{Y}_M) = E_1V_2E_3\left(\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{D_h}{d_h}\sum_{j=1}^{d_h}\sum_{k=1}^{T}\frac{M_{hijk}}{m_{hijk}}\sum_{l=1}^{m_{hijk}}y_{hijkl}\right)$$

$$E_1V_2E_3(\hat{Y}_M) = E_1V_2\left(\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{D_h}{d_h}\sum_{j=1}^{d_h}Y_{hij}\right)$$

$$= E_1\left(\sum_{h=1}^{L}\frac{N_h^2}{n_h^2}\sum_{i=1}^{n_h}V_2\left(\frac{D_h}{d_h}\sum_{j=1}^{d_h}Y_{hij}\right)\right)$$

$$= E_1\left(\sum_{h=1}^{L}\frac{N_h^2}{n_h^2}\sum_{i=1}^{n_h}D_h^2V_2\left(\bar{Y}_{hi}\right)\right)$$

$$= E_1\left(\sum_{h=1}^{L}\frac{N_h^2}{n_h^2}\sum_{i=1}^{n_h}D_h^2\left(\frac{1}{d_h}-\frac{1}{D_h}\right)S_{hi}^2\right);$$

where

$$S_{hi}^2 = \frac{1}{D_h-1}\sum_{i=1}^{D_h}\left(Y_{hij}-\bar{Y}_{hi}\right)^2 \qquad \text{----(5.17)}$$

since the second stage selection also is as per SRSWOR.

Hence,

$$E_1V_2E_3(\hat{Y}_M) = \sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{N_h}D_h^2\left(\frac{1}{d_h}-\frac{1}{D_h}\right)S_{hi}^2 \qquad \text{-----(5.18)}$$

To evaluate the third term, we use the following procedure.

Let $\bar{y}'_{st}$ denote the mean of a post-stratified sample of size n, drawn from a population with N units with strata weights denoted as $W_g = \frac{N_g}{N}$,

$g=1,2,\ldots,L$. Then $\bar{y}'_{st}=\sum_{g=1}^{L}W_g\bar{y}_g$ is the post-stratified estimator of the population mean.

Then

$$V(\bar{y}'_{st}) \qquad = EV(\bar{y}'_{st}/n_g)+VE(\bar{y}'_{st}/n_g)$$

$$= EV(\bar{y}'_{st}/n_g),$$

since $E(\bar{y}'_{st}/n_g)$ is a constant.

$$V(\bar{y}'_{st}) \qquad = E\left(\sum_{g=1}^{L}W_g^2\left(\frac{1}{n_g}-\frac{1}{N_g}\right)S_g^2\right)$$

$$=\sum_{g=1}^{L}\left(E\left(\frac{1}{n_g}\right)-\frac{1}{N_g}\right)W_g^2S_g^2 \qquad \text{-----(5.19)}$$

One approximation for $E\left(\frac{1}{n_g}\right)$ given by Sukhatme *et al.* (1997) is

$$E\left(\frac{1}{n_g}\right) \qquad =\frac{1}{nW_g}+\frac{1-W_g}{n^2W_g^2}$$

A still better estimator of $E\left(\frac{1}{n_g}\right)$ can be obtained as follows. The number of observations falling into the g^{th} stratum, n_g, on post-stratification, can be regarded as a hyper-geometric random variable with parameters (N,N_g,n), so that

$$E(n_g)=n\frac{N_g}{N}$$

$$V(n_g)=\left(\frac{N-n}{N-1}\right)\left(n\frac{N_g}{N}\right)\left(1-\frac{N_g}{N}\right)$$

Denoting $\frac{n_g - E(n_g)}{E(n_g)} = \Delta_g$,

We get

$$\frac{1}{n_g}=\frac{1}{n(1+\Delta_g)}$$

$$=\frac{1}{n}\left(1-\Delta_g+\Delta_g^2+\ldots\right)$$

$\cong\frac{1}{n}\left(1-\Delta_g+\Delta_g^2\right)$ by neglecting terms in higher powers of Δ_g than the second.

Also $E(\Delta_g)=0$

$$V(\Delta_g)=\frac{V(n_g)}{\left[E(n_g)\right]^2}$$

Thus we get

$$E\left(\frac{1}{n_g}\right) \quad = \frac{1}{E(n_g)}\left(1+\frac{V(n_g)}{\left[E(n_g)\right]^2}\right) \qquad \text{-----(5.20)}$$

Using (5.20) we get the expression for $V(\bar{y}'_{st})$ in (5.19) as

$$V(\bar{y}'_{st}) \quad =\sum_{g=1}^{L}\left(\frac{1}{E(n_g)}\left(1+\frac{V(n_g)}{\left[E(n_g)\right]^2}\right)-\frac{1}{N_g}\right)W_g^2S_g^2 \qquad \text{-----(5.21)}$$

In our present case $n_g = m_{hijk}, N_g = M_{hijk}$, $n = m_{hij}$ and $N = M_{hij}$ so that

$$E(m_{hijk}) \quad = m_{hij}\frac{M_{hijk}}{M_{hij}} \qquad \text{-----(5.22)}$$

$$V(m_{hijk}) = \left(\frac{M_{hij} - m_{hij}}{M_{hij} - 1}\right)\left(\frac{m_{hij} M_{hijk}}{M_{hij}}\right)\left(1 - \frac{M_{hijk}}{M_{hij}}\right) \qquad \text{-----(5.23)}$$

Now the third term of (5.15) is

$$E_1 E_2 V_3(\hat{Y}_M) = E_1 E_2 V_3\left(\sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{D_h}{d_h} \sum_{j=1}^{d_h} \sum_{k=1}^{T} \frac{M_{hijk}}{m_{hijk}} \sum_{l=1}^{m_{hijk}} y_{hijkl}\right)$$

$$= E_1 E_2 \sum_{h=1}^{L} \frac{N_h^2}{n_h^2} \sum_{i=1}^{n_h} \frac{D_h^2}{d_h^2} \sum_{j=1}^{d_h} V_3\left(M_{hijk} \sum_{k=1}^{T} W_{hijk} \bar{y}_{hijk}\right)$$

$$= E_1 E_2 \sum_{h=1}^{L} \frac{N_h^2}{n_h^2} \sum_{i=1}^{n_h} \frac{D_h^2}{d_h^2} \sum_{j=1}^{d_h} M_{hijk}^2 V_3\left(\hat{\bar{Y}}_{hij}\right)$$

Treating $\hat{\bar{Y}}_{hij}$ as $\bar{y}'_{st}$, its variance is given by equation (5.21). Thus we get the above as

$$E_1 E_2 V_3(\hat{Y}_M) = E_1 E_2 \left(\sum_{h=1}^{L} \frac{N_h^2}{n_h^2} \sum_{i=1}^{n_h} \frac{D_h^2}{d_h^2} \sum_{j=1}^{d_h} M_{hij}^2 \sum_{k=1}^{T} \frac{M_{hij}}{m_{hij} M_{hijk}} \left[\left[1 + \frac{\left(\frac{M_{hij} - m_{hij}}{M_{hij} - 1}\right)\left(\frac{m_{hij} M_{hijk}}{M_{hij}}\right)\left(1 - \frac{M_{hijk}}{M_{hij}}\right)}{\left(\frac{m_{hij} M_{hijk}}{M_{hij}}\right)^2}\right] - \frac{1}{M_{hijk}}\right] W_{hijk}^2 S_{hijk}^2\right)$$

$$= E_1 E_2 \left(\sum_{h=1}^{L} \frac{N_h^2}{n_h^2} \sum_{i=1}^{n_h} \frac{D_h^2}{d_h^2} \left[\sum_{j=1}^{d_h} \left(\frac{1}{m_{hij}} - \frac{1}{M_{hij}}\right) \sum_{k=1}^{T} M_{hijk}^2 S_{hijk}^2 + \frac{M_{hij}}{m_{hij}} \left(\frac{M_{hij} - m_{hij}}{M_{hij} - 1}\right) \sum_{k=1}^{T} \left(1 - \frac{M_{hijk}}{M_{hij}}\right) S_{hijk}^2\right]\right)$$

where

$$S_{hijk}^2 = \frac{1}{M_{hijk}-1}\sum_{l=1}^{M_{hijk}}\left(Y_{hijkl}-\bar{Y}_{hijk}\right)^2 \qquad \text{-----(5.24)}$$

$$E_1E_2V_3(\hat{Y}_M) = E_1\left(\sum_{h=1}^{L}\frac{N_h^2}{n_h^2}\sum_{i=1}^{n_h}\frac{D_h}{d_h}\sum_{j=1}^{D_h}\left[\left(\frac{1}{m_{hij}}-\frac{1}{M_{hij}}\right)\sum_{k=1}^{T}M_{hijk}^2S_{hijk}^2 + \frac{M_{hij}}{m_{hij}}\left(\frac{M_{hij}-m_{hij}}{M_{hij}-1}\right)\sum_{k=1}^{T}\left(1-\frac{M_{hijk}}{M_{hij}}\right)S_{hijk}^2\right]\right)$$

$$= \sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{D_h}{d_h}\sum_{j=1}^{D_h}\left[\left(\frac{1}{m_{hij}}-\frac{1}{M_{hij}}\right)\sum_{k=1}^{T}M_{hijk}^2S_{hijk}^2 + \frac{M_{hij}}{m_{hij}}\left(\frac{M_{hij}-m_{hij}}{M_{hij}-1}\right)\sum_{k=1}^{T}\left(1-\frac{M_{hijk}}{M_{hij}}\right)S_{hijk}^2\right] \qquad \text{-----(5.25)}$$

Now substituting (5.16), (5.18) and (5.25) in (5.15), we get

$$V(\hat{Y}_M) = \sum_{h=1}^{L}\left\{N_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)S_h^2 + \frac{N_h}{n_h}\sum_{i=1}^{N_h}D_h^2\left(\frac{1}{d_h}-\frac{1}{D_h}\right)S_{hi}^2 + \frac{N_h}{n_h}\sum_{i=1}^{N_h}\frac{D_h}{d_h}\sum_{j=1}^{D_h}\left[\left(\frac{1}{m_{hij}}-\frac{1}{M_{hij}}\right)\sum_{k=1}^{T}M_{hijk}S_{hijk}^2 + \frac{M_{hij}}{m_{hij}}\left(\frac{M_{hij}-m_{hij}}{M_{hij}-1}\right)\sum_{k=1}^{T}\left(1-\frac{M_{hijk}}{M_{hij}}\right)S_{hijk}^2\right]\right\} \qquad \text{-----(5.26)}$$

Note that the variance of the estimator contains four components, the first and second terms in (5.26) representing the variations due to first stage strata and second stage respectively while the third and fourth terms represent the variation due to the third stage (post-stratification). The estimate of variance $\hat{V}(\hat{Y}_M)$ is given by substituting S_h^2, S_{hi}^2 and S_{hijk}^2 with their corresponding sample estimators s_h^2, s_{hi}^2 and s_{hijk}^2.

$$\hat{V}(\hat{Y}_M)=\sum_{h=1}^{L}\left\{N_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)s_h^2+\frac{N_h}{n_h}\sum_{i=1}^{N_h}D_h^2\left(\frac{1}{d_h}-\frac{1}{D_h}\right)s_{hi}^2\right.$$

$$+\frac{N_h}{n_h}\sum_{i=1}^{N_h}\frac{D_h}{d_h}\sum_{j=1}^{D_h}\left[\left(\frac{1}{m_{hij}}-\frac{1}{M_{hij}}\right)\sum_{k=1}^{T}M_{hijk}s_{hijk}^2\right.$$

$$\left.\left.+\frac{M_{hij}}{m_{hij}}\left(\frac{M_{hij}-m_{hij}}{M_{hij}-1}\right)\sum_{k=1}^{T}\left(1-\frac{M_{hijk}}{M_{hij}}\right)s_{hijk}^2\right]\right\} \qquad \text{-----(5.27)}$$

where

$$s_h^2 = \frac{1}{n_h-1}\sum_{i=1}^{n_h}\left(Y_{hi}-\bar{Y}_h\right)^2 \qquad \text{-----(5.28)}$$

$$s_{hi}^2 = \frac{1}{d_h-1}\sum_{i=1}^{d_h}\left(Y_{hij}-\bar{Y}_{hi}\right)^2 \qquad \text{-----(5.29)}$$

$$s_{hijk}^2 = \frac{1}{m_{hijk}-1}\sum_{i=1}^{m_{hijk}}\left(Y_{hijkl}-\bar{Y}_{hijk}\right)^2 \qquad \text{-----(5.30)}$$

State Level Estimate

The first estimator proposed above (5.2) is for a single centre zone and the second (5.11) is for a multi-centre zone. Let Z_1 and Z_2 denote the number of single and multi-centre zones in the state and $\hat{Y}_z$ and $\hat{Y}_{Mz}$ denote the post-stratified estimators as given by (5.2) and (5.11) respectively. Then the estimate of the total landings for a month for all the zones in the state is given by

$$\hat{Y}_s = \sum_{z=1}^{Z_1}\hat{Y}_z+\sum_{z=1}^{Z_2}\hat{Y}_{Mz}$$

and the estimate of the variance of the state level estimator for the month is given by

$$V(\hat{Y}_s) = \sum_{z=1}^{Z_1} \hat{V}(\hat{Y}_z) + \sum_{z=1}^{Z_2} \hat{V}(\hat{Y}_{Mz})$$

where $\hat{V}(\hat{Y}_z)$ and $\hat{V}(\hat{Y}_{Mz})$ are given by (5.7) and (5.27) respectively.

5.3 Conclusion

One significant aspect of the new designs proposed in this chapter is their inherent characteristic to ensure higher sampling fraction. In the first design for the single centre zone, by increasing the sample size n appropriately any desired sampling fraction can be achieved. In the case of the second estimator proposed for multi-centre zones the sampling fraction achieved is $f' = \frac{\sum_{h=1}^{L} n_h d_h}{\sum_{h=1}^{L} N_h D_h}$. The sampling fraction that can be achieved in an equivalent existing design is $f = \frac{\sum_{h=1}^{L} n_h}{\sum_{h=1}^{L} N_h D_h}$. Suppose we fix $d_h = c$ for all h. Then $f' = f.c$. Thus the new method increases the sampling fraction c-fold. For c=1, the sampling fraction reduces to that of the existing design. Another, advantage is that due to adopting post-stratification, the estimate of variance takes into account the variation due to gears together with variation between days which were not accounted in the existing design. Numerical illustration reveals that the first design leads to a more efficient estimator. The second method could not be illustrated to establish its effectiveness for want of adequate data in any of the existing multi-centre zones. However, the above listed specific advantages provide a sufficient proof to establish the effectiveness of the new design. The two new designs introduced in this chapter rectify most of the limitations of the existing design described in chapter 2.

CHAPTER VI

Estimation of Fish Landings – PPS Sampling Design

6.1 Introduction

With an intention to overcome the drawbacks of the already existing sampling design, we have introduced two new multi-stage designs in chapter 5, one for single centre zones and the other for multi-centre zones. Both these designs mainly aim at increasing the sampling fraction and estimating the variance of the estimator more scientifically. However, it can be seen that a slight increase in the sampling fraction causes exponential increase in the cost of the survey. In order to reduce the escalation of cost of the survey as well as simplifying the sampling procedure in the case of sampling from a multi-centre zone, a new two stage design which is structurally and operationally simple than the three stage stratified design described in section 5.3 is introduced in this chapter. Different procedures of estimating the optimum sample size in the designs are also described in this chapter.

Since the sampling fraction achieved in the existing sampling design is very low, any modification of the design must increase the sampling fraction. To reduce the high cost of the survey due to increasing the sampling fraction without affecting the reliability of the estimator in the new design, we propose to utilize the past information to the maximum extent. There are several methods of utilizing the past information at the design level and at the estimation level. The most popular method of utilizing the past information at the design level is by adopting the probability proportional to size (PPS) method of sample selection where the probabilities of selection are fixed on the basis of past information. Based on the trend analysis made in chapter 3 we have seen that the total landings are highly irregular over the months and over the years, however, it has been noted that some of the zones indicate an approximately decreasing tendency towards the beginning periods of the monsoon season or a slightly increasing tendency towards the later half of the year etc. Collection of fish landings data being a continuous process, the estimate of total catch will be available for all months, which is utilized for PPS selection. Hence, the proposed new design is to adopt a PPS method of selecting the PSUs (landing centres) where the probabilities of selection are taken as proportional to the catch at the landing centre for the month in the previous year. Since the analysis of variance made in chapter 4 revealed that the day-wise variation in the fish landings data is not significant quiet often, the modified design does not account the day-wise variation explicitly. Another modification of the proposed design is craft-wise estimation of the landings instead of the hitherto used gear-wise estimates. The last modification is proposed since, as of late, most of the boats use multiple gears and hence it is not possible to attribute the catches in a boat to a particular type of gear alone. At present the fishing crafts are broadly classified into three as mechanised, motorised and non-mechanised. A still further homogeneous division of the crafts can be made in terms of their size, fishing power and capacity.

The new two stage PPS design is described in section 6.2. Two types of estimators - craft-wise and day-wise and their variances are developed in this section. Methods of estimating the optimum sample size in the case of the different designs are described in detail in section 6.3. The chapter ends with a brief conclusion.

6.1 Two Stage PPS Design

Consider a multi-centre zone with N landing centres, let X_i, $(i = 1,2,...,N)$ denotes the estimated total landings at the i^{th} landing centre for the month during the previous year, $p_i = \frac{X_i}{X}$, with $X = \sum_{i=1}^{N} X_i$. Adopting PPSWR, n landing centres are selected with p_i as the probability of selecting the i^{th} landing centre. These centres are evenly distributed among the available field staff and are directed to visit the centres one by one for a single day each during the month. Because of using PPSWR, there is the possibility of getting the same landing centre more than once in the sample. In such cases the repeated occurrence of a centre is treated independently for field visit. On the day of visit, fishing boats landed (SSU) selected for observation at intervals of 15 to 20 minutes. As in the new designs proposed in chapter 5, here also 24 hr duration of the day of visit is regarded as a single unit. Count and catch are recorded together. Night landings are accounted only for count by enquiry on the following day. During the first half day of the visit priority is given to boats of distinct crafts type for recording catch. In the following half day of visit the priority has been given to get adequate representation for each craft type depending on the total number of boats of each craft landed including night landings. Since the number of distinct craft types may be far less than the number of gear types and are easily identifiable, it is very easy to ensure adequate representation for the boats of each craft type in the sample. To simplify the estimation process the SSU selection is regarded as according to SRSWOR.

Let there be T_i types of crafts at the i^{th} landing centre and M_{ik} and m_{ik} denote the total and sampled number of fishing boats respectively of k^{th} type craft landed at the i^{th} centre.

Let y_{ikl} be the observed landings (sum of all species) by the l^{th} boat of the k^{th} craft at the i^{th} landing centre ; $l = 1,2,...,m_{ik}$, $i = 1,2,...,n$.

As usual, the upper case symbols represent the population characteristics and the symbols with an upper bar denote the corresponding means.

Then the average quantity of fish landed by the k^{th} craft type at the i^{th} centre is given by

$$\bar{y}_{ik.} = \frac{\sum_{l=1}^{m_{ik}} y_{ikl}}{m_{ik}} \qquad \text{-----(6.1)}$$

6.2.1 PPS Estimator of Landings per Craft

An estimator for the total quantity of fish landed by the k^{th} craft ($\hat{Y}_k$) is given by

$$\hat{Y}_k = \frac{1}{n}\sum_{i=1}^{n}\frac{M_{ik}\bar{y}_{ik.}}{p_i} \qquad \text{-----(6.2)}$$

where

$$p_i = \frac{X_i}{X} \qquad \text{-----(6.3)}$$

X_i being the estimate of landings for the month in the previous year for the zone. $i = 1,2,...,N$ and $X = \sum_{i=1}^{N} X_i$

Then

$$E(\hat{Y}_k) = E_1 E_2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{ik}\bar{y}_{ik.}}{p_i}\right)$$

where E_1 and E_2 are expectations with respect to the selection of PSU and SSU respectively. Then,

$$E(\hat{Y}_k) = E_1\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}E_2(M_{ik}\bar{y}_{ik.})\right)$$

$$= E_1\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}(M_{ik}\bar{Y}_{ik.})\right)$$

since, $\bar{y}_{ik.}$ is average landings of m_{ik} fishing boats selected out of M_{ik} by SRS, it is a unbiased estimator for $\bar{Y}_{ik.}$.

$$E(\hat{Y}_k) = \left(\sum_{i=1}^{N}\frac{p_i}{p_i}M_{ik}\bar{Y}_{ik.}\right)$$

$$= \left(\sum_{i=1}^{N}M_{ik}\bar{Y}_{ik.}\right)$$

$$E(\hat{Y}_k) \quad = Y_k \qquad \text{-----(6.4)}$$

Therefore $\hat{Y}_k$ is an unbiased estimator of Y_k, where Y_k is the total fish landings by k^{th} gear.

Variance of the craft estimator

The variance of $\hat{Y}_k$ under a two stage sampling (Desraj, 1971) is given by

$$V(\hat{Y}_k) \qquad = E_1V_2(\hat{Y}_k) + V_1E_2(\hat{Y}_k) \qquad \text{-----(6.5)}$$

where the subscripts 1 and 2 indicate that the expectation and variance are taken with respect to the first and second stages of sampling respectively.

Now the first term of (6.5) is given by

$$E_1V_2(\hat{Y}_k) \qquad = E_1V_2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{ik}\bar{y}_{ik.}}{p_i}\right)$$

$$= E_1\left(\frac{1}{n^2}\sum_{i=1}^{n}\frac{M_{ik}^2}{p_i^2}V_2(\bar{y}_{ik.})\right)$$

$$= E_1\left(\frac{1}{n^2}\sum_{i=1}^{n}\frac{M_{ik}^2}{p_i^2}\left(\frac{1}{m_{ik}}-\frac{1}{M_{ik}}\right)S_{ik}^2\right),$$

where

$$S_{ik}^2 = \frac{1}{M_{ik}-1}\sum_{l=1}^{M_{ik}}(y_{ikl.}-\bar{Y}_{ik.})^2 \qquad \text{-----(6.6)}$$

Therefore

$$E_1V_2(\hat{Y}_k) \quad = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i=1}^{N}\frac{p_i M_{ik}^2}{p_i^2}\left(\frac{1}{m_{ik}}-\frac{1}{M_{ik}}\right)S_{ik}^2$$

$$E_1V_2(\hat{Y}_k) \quad = \frac{1}{n}\sum_{i=1}^{N}\frac{M_{ik}^2}{p_i}\left(\frac{1}{m_{ik}}-\frac{1}{M_{ik}}\right)S_{ik}^2 \qquad \text{-----(6.7)}$$

Similarly, the second term of (6.5) is

$$V_1E_2(\hat{Y}_k) \quad = V_1E_2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{ik}\bar{y}_{ik.}}{p_i}\right)$$

$$= V_1\left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{ik}\bar{Y}_{ik.}}{p_i}\right)$$

Under PPS scheme, the variance of $(\hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}$ is given by $\frac{1}{n}\sum_{i=1}^{N}p_i\left(\frac{Y_i}{p_i}-Y\right)^2$. In the present case, we have $y_i = M_{ik}\bar{Y}_{ik.}$ and hence,

$$V_1E_2(\hat{Y}_k) \quad = \frac{1}{n}\sum_{i=1}^{N}p_i\left(\frac{Y_{ik.}}{p_i}-Y_k\right)^2 \qquad \text{-----(6.8)}$$

substituting (6.7) and (6.8) in the equation (6.5), we get

$$V(\hat{Y}_k) \quad = \frac{1}{n}\sum_{i=1}^{N}\frac{M_{ik}^2}{p_i}\left(\frac{1}{m_{ik}}-\frac{1}{M_{ik}}\right)S_{ik}^2 + \frac{1}{n}\sum_{i=1}^{N}p_i\left(\frac{Y_{ik.}}{p_i}-Y_k\right)^2 \qquad \text{-----(6.9)}$$

An unbiased estimator of $\sum_{i=1}^{N}p_i\left(\frac{Y_{ik.}}{p_i}-Y_k\right)^2$ is given by Desraj (1971) as $\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{Y_{ik.}}{p_i}-\hat{Y}_k\right)^2$ and $s_{ik}^2 = \frac{1}{m_{ik}-1}\sum_{l=1}^{m_{ik}}(y_{ikl.}-\bar{y}_{ik.})^2$ is unbiased for S_{ik}^2, then we get an unbiased estimator of $V(\hat{Y}_k)$ as

$$\hat{V}(\hat{Y}_k) = \frac{1}{n}\sum_{i=1}^{n}\frac{M_{ik}^2}{p_i}\left(\frac{1}{m_{ik}} - \frac{1}{M_{ik}}\right)s_{ik}^2 + \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{Y_{ik.}}{p_i} - \hat{Y}_k\right)^2 \qquad \text{----(6.10)}$$

for each k.

6.2.2 PPS Estimator of Landings per Day

Let T denote the number of distinct crafts landed in a zone over a month. An estimator for the total quantity of fish landed on a day in the zone ($\hat{Y}_{ppz}$) is given by

$$\hat{Y}_{ppz} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{Y}_{i.}}{p_i}, \qquad \text{-----(6.11)}$$

where $\hat{Y}_{i.} = \sum_{k=1}^{T} M_{ik}\bar{y}_{ik.}$

The above estimator is unbiased for the total catch/day at the zone,

For,

$$E(\hat{Y}_{ppz}) = E_1E_2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{Y}_{i.}}{p_i}\right)$$

$$= E_1E_2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}\sum_{k=1}^{T}M_{ik}\bar{y}_{ik.}\right)$$

$$= E_1\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}\sum_{k=1}^{T}E_2(M_{ik}\bar{y}_{ik.})\right)$$

$$= E_1\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}\sum_{k=1}^{T}(M_{ik}\bar{Y}_{ik.})\right)$$

$$E(\hat{Y}_{ppz}) = E_1\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}(\hat{Y}_{i.})\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{i=1}^{N}p_i\left(\frac{\hat{Y}_{i.}}{p_i}\right)$$

$$= Y \qquad \text{-----(6.12)}$$

Therefore $\hat{Y}_{ppz}$ is an unbiased estimator of Y, where Y is the total fish landings in a day.

Variance of the day estimator

$$V(\hat{Y}_{ppz}) = E_1V_2(\hat{Y}_{ppz}) + V_1E_2(\hat{Y}_{ppz}) \qquad \text{-----(6.13)}$$

Now taking the first term of (6.13),

$$E_1V_2(\hat{Y}_{ppz}) = E_1V_2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{Y}_{i.}}{p_i}\right)$$

$$= E_1V_2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}\sum_{k=1}^{T}M_{ik}\bar{y}_{ik.}\right)$$

$$= E_1\left(\frac{1}{n^2}\sum_{i=1}^{n}\frac{M_{ik}^2}{p_i^2}\sum_{k=1}^{T}V_2(\bar{y}_{ik.})\right)$$

$$= E_1\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{T}\frac{M_{ik}^2}{p_i^2}\left(\frac{1}{m_{ik}} - \frac{1}{M_{ik}}\right)S_{ik}^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{N}\sum_{k=1}^{T}\frac{p_iM_{ik}^2}{p_i^2}\left(\frac{1}{m_{ik}} - \frac{1}{M_{ik}}\right)S_{ik}^2 \qquad \text{-----(6.14)}$$

similarly,

$$V_1E_2(\hat{Y}_{ppz}) \quad = V_1E_2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{Y}_{i.}}{p_i}\right)$$

$$= V_1\left(\frac{1}{n}\sum_{i=1}^{n}\frac{Y_{i.}}{p_i}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{N}p_i\left(\frac{Y_{i.}}{p_i} - Y\right)^2 \qquad \text{-----(6.15)}$$

substituting (6.14) and (6.15) in the equation (6.13), we get

$$V(\hat{Y}_{ppz}) \quad = \frac{1}{n}\sum_{i=1}^{N}\sum_{k=1}^{T}\frac{M_{ik}^2}{p_i}\left(\frac{1}{m_{ik}} - \frac{1}{M_{ik}}\right)S_{ik}^2 + \frac{1}{n}\sum_{i=1}^{N}p_i\left(\frac{Y_{i.}}{p_i} - Y\right)^2, \qquad \text{-----(6.16)}$$

where S_{ik}^2 is given by (6.6). Replacing the population factors by their unbiased estimators as in the case of (6.9) we get the unbiased estimator of the variance of the day estimator as

$$\hat{V}(\hat{Y}_{ppz}) \quad = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{T}\frac{M_{ik}^2}{p_i}\left(\frac{1}{m_{ik}} - \frac{1}{M_{ik}}\right)s_{ik}^2 + \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{Y_{i.}}{p_i} - \hat{Y}\right)^2 \qquad \text{----(6.17)}$$

Thus we have provided two estimators; one for estimating the craft-wise landings/day and another for the day-wise landings. Note that both the estimators give the average landings per day for the zone. Hence, the monthly estimate can be obtained by simply multiplying the day average by the number of fishing days in the month. The variances of the monthly estimates also can be obtained in the same manner. It may be noted that the day estimator (6.11) is the sum of the craft estimators (6.2) where the summation is made over the crafts.

$$\hat{Y} = \sum_{k}^{T} \hat{Y}_k \qquad \text{----(6.18)}$$

However, the variance of the day estimator is not the same as the sum of the variances of the craft estimators as indicated below.

The variance of the craft estimator is given by (6.9)

$$V(\hat{Y}_k) = \frac{1}{n}\sum_{i=1}^{N}\frac{M_{ik}^2}{p_i}\left(\frac{1}{m_{ik}}-\frac{1}{M_{ik}}\right)S_{ik}^2 + \frac{1}{n}\sum_{i=1}^{N}p_i\left(\frac{Y_{ik.}}{p_i}-Y_k\right)^2$$

Now summing the above with respect to k, we get

$$\sum_{k}^{T}V(\hat{Y}_k) = \frac{1}{n}\sum_{k=1}^{T}\sum_{i=1}^{N}\frac{M_{ik}^2}{p_i}\left(\frac{1}{m_{ik}}-\frac{1}{M_{ik}}\right)S_{ik}^2 + \frac{1}{n}\sum_{k=1}^{T}\sum_{i=1}^{N}p_i\left(\frac{Y_{ik.}}{p_i}-Y_k\right)^2.$$

Now

$$\frac{1}{n}\sum_{k=1}^{T}\sum_{i=1}^{N}p_i\left(\frac{Y_{ik.}}{p_i}-Y_k\right)^2 = \sum_{i=1}^{N}p_i\sum_{k=1}^{T}\left(\frac{Y_{ik.}}{p_i}-Y_k\right)^2$$

$$= \sum_{i=1}^{N}p_i\left\{\left[\sum_{k=1}^{T}\left(\frac{Y_{ik.}}{p_i}-Y_k\right)\right]^2 - \sum_{k\neq l}^{T}\left(\frac{Y_{ik.}}{p_i}-Y_k\right)\left(\frac{Y_{il.}}{p_i}-Y_l\right)\right\}$$

$$= \sum_{i=1}^{N}p_i\left\{\left(\frac{Y_{i.}}{p_i}-Y\right)^2 - \sum_{k\neq l}^{T}\left(\frac{Y_{ik.}}{p_i}-Y_k\right)\left(\frac{Y_{il.}}{p_i}-Y_l\right)\right\} \qquad \text{----(6.19)}$$

Using (6.19) we get

$$\sum_{k}^{T}V(\hat{Y}_k) = \frac{1}{n}\sum_{k}\sum_{i=1}^{N}\frac{M_{ik}^2}{p_i}\left(\frac{1}{m_{ik}}-\frac{1}{M_{ik}}\right)S_{ik}^2$$

$$+\frac{1}{n}\sum_{i=1}^{N}p_i\left\{\left(\frac{Y_{i.}}{p_i}-Y\right)^2 - \sum_{k\neq l}^{T}\left(\frac{Y_{ik..}}{p_i}-Y_k\right)\left(\frac{Y_{il..}}{p_i}-Y_l\right)\right\},$$

which can be written as

$$\sum_{k}^{T} V(\hat{Y}_k) = V(\hat{Y}) - \frac{1}{n}\sum_{i=1}^{n} p_i \sum_{k\neq l}^{T}\left(\frac{Y_{ik.}}{p_i} - Y_k\right)\left(\frac{Y_{il.}}{p_i} - Y_l\right) \qquad \text{----(6.20)}$$

(6.18) and (6.20) indicate that though the day estimator can be obtained as the direct sum of the craft wise estimators, the variance of the day estimator is to be computed separately instead of simply summing the variances of the craft estimators.

6.2.3 Sampling fraction

In the development of the PPS estimator we have assumed that a landing centre selected to the sample is observed only for a single day. Because of adopting With Replacement sampling it is likely that the same centre is observed for more than one day. However, if it is required to increase the sampling fraction to any desired level, the same can be ensured by a slight modification of the field visit as described below without causing any significant change in the estimation process.

Instead of collecting data from a selected centre for only one day, the field staff may observe the centre for *d* days continuously. On each repeated day, the data is collected in the same manner as on the first day of visit. Then the average of the *d* days observations is taken as the data corresponding to the landing centre. Thus if y_{iklj} denote the landings for the j^{th} day, M_{ikj} and m_{ikj} are the number of boats landed and observed on the j^{th} day, then redefine the symbols of section 6.2.2 as

$$y_{ikl} = \frac{\sum_{j=1}^{d} y_{iklj}}{d}, \; M_{ik} = \frac{\sum_{j=1}^{d} M_{ikj}}{d} \text{ and } m_{ik} = \frac{\sum_{j=1}^{d} m_{ikj}}{d}.$$

Since the ANOVA performed in chapter 4 reveals that the day-wise variation of the fish landing data is negligible, the above averaging will not cause any significant change in the total variance of the estimator and hence the same estimators (6.2) and (6.11) and their variance expressions can be used with the above redefined data values. It may be noted that the above indicated modification is highly flexible in the sense that the value of d can be chosen independently for each centre. The choice may be based on the intensity of fishing operations being noted at the centre on the first day of visit. This facility though not essential is an additional feature of the design allowing adjustments in the deployment of field staff to the centres at which more intensive fishing operations are occurring in a particular month.

6.2.4 Estimation of PPS Selection Probabilities

For applying the PPS design, we require the values of p_i for all centres in the zone, $p_i = \frac{X_i}{X}$, X_i being the total landings at the i^{th} centre for the month and $X = \sum_{i=1}^{N} X_i$. If it can be assumed that the number of fishing days are the same for all the centres in the zone then the value of p_i can be obtained by replacing X_i by the total landings per day at the centre. Let n' denote the number of distinct landing centres observed at the zone and $X' = X_1 + X_2 + ... + X_{n'}$. Then the average landings per day for the $N-n'$ unobserved landing centres is given by $\bar{X}=\dfrac{X-X'}{N-n'}$. Now

$X=\hat{Y}_{ppz}$ and

$X'=\hat{Y}_i$

$=\sum_k M_{ik}\bar{y}_{ik.}$, where $\bar{y}_{ik.}$ is given by (6.1).

Then the estimatof of p_i is given by

$p_i=\dfrac{X_i}{X}$, for $i=1,2,...,n'$

$=\dfrac{\bar{X}}{X}$, for $i=n'+1,...,N$

It may be noted that as per the above procedure the estimate of p_i for all unobserved centres are the same. To get a more precise estimator of p_i, other option is to conduct a rapid census for collecting the data on total landings in all centres. If we get information on the total boats M_{ik} operated at each of the unobserved centres, then the above estimator can be modified as

$$p_i'=\frac{X_i'}{X}, \text{ where } X_i'=\sum_k \frac{M_{ik}}{N}\hat{Y}_k \ , \ i=n'+1,...,N.$$

where $\hat{Y}_k$ is the k^{th} craft estimator given by (6.2).

6.3 Estimation of Optimum Sample Size

Among the different criteria of deriving the optimum sample size in a random survey, the two most popular criteria are

(a) to obtain the estimate of the population total (or mean) within a specified margin of error and

(b) to minimise the variance of the estimator subject to some given constraint such as fixed total cost.

Under the criteria (a) if $\hat{Y}_N$ denote the unbiased estimator of the population total Y_N based on a random sample of size n drawn from a population of size N and δ is the maximum permissible margin of error in the estimate (δ may be specified as some percentage of $\hat{Y}_N$). Then the sample size n is estimated using the condition

$$P\left\{\left|\hat{Y}_N - Y_N\right| \le \delta \right\} = 1-\alpha \qquad \text{-----(6.21)}$$

where $1-\alpha$ is the confidence level.

(6.21) can be rewritten as

$$P\left\{\left|\frac{\hat{Y}_N - Y_N}{SE(\hat{Y}_N)}\right| \le \frac{\delta}{SE(\hat{Y}_N)}\right\} = 1-\alpha \qquad \text{-----(6.22)}$$

Assuming that the data follow a normal law, $\frac{\hat{Y}_N - Y_N}{SE(\hat{Y}_N)}$ can be taken as a student's t-variate with (n-1) df. Thus if t' denote the value of t such that

$$P\{t_{n-1} \le t'\} = 1-\alpha/2 \qquad \text{-----(6.23)}$$

Then (6.22) and (6.23) will imply $t' = \frac{\delta}{SE(\hat{Y}_N)}$

$$\Rightarrow V(\hat{Y}_N) = \left(\frac{\delta}{t'}\right)^2 \qquad \text{-----(6.24)}$$

Solving (6.24) we get the value of n satisfying criteria (6.21).

In the case of SRS, (6.24) gives the optimum value of n as

$$n = \frac{1}{\frac{1}{N} + \frac{1}{N^2}\left(\frac{\delta}{t'}\right)^2 \frac{1}{S^2}}, \qquad \text{-----(6.25)}$$

where S^2 being the population mean square.

Note that while using (6.25) for the value of t', we require advance knowledge of n. In the absence of any knowledge about the approximate size of the sample, the value of t' may be replaced by the corresponding value of standard normal. In case the value of n is suspected to be very small (less than 25), then (6.25) may be used assuming an initial value, $n_0 = 25$. If the estimate of n is near 25 then, admit the value as such else if the estimate of n is too small then decrease the value of n_0 slightly and repeat the formula until the estimate somewhat agree with the approximately selected value. (It may be noted that decreasing the degrees of freedom increases the value of t' which correspondingly will cause an increase in the estimate of n.)

6.3.1 Optimum sample size for multi-centre zone in the existing design

In the existing sampling design, the estimation of variance in the case of the multi-centre zones are made regarding the survey as a single stage stratified sampling. Using the criteria (a) the overall sample size for a multi-centre zone can be fixed as follows.

Let $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$ denote the unbiased estimate of the population mean based on a stratified random sampling (with usual notations). The estimate of the population total is $\hat{Y}_N = N\bar{y}_{st}$ with variance, $V(\hat{Y}_N) = N^2 \sum_{h=1}^{L} W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_h^2$. Following the same assumptions as in the case of SRS, the equations equivalent to (6.24) is

$$N^2 \sum_{h=1}^{L} W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_h^2 = \left(\frac{\delta}{t'} \right)^2 \qquad \text{-----(6.25)}$$

Note that the sample size n is not explicitly present in (6.25) causing difficulty to estimate it. Hence, we introduce the sample analogue of the stratum weight as

$$w_h = \frac{n_h}{n} \text{ so that } n_h = n\, w_h . \qquad \text{-----(6.26)}$$

Using (6.26) in (6.25), we get

$$N^2 \sum_{h=1}^{L} W_h^2 \left(\frac{1}{n\, w_h} - \frac{1}{N_h} \right) S_h^2 = \left(\frac{\delta}{t'} \right)^2$$

which gives,

$$\frac{1}{n} \left[N^2 \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{w_h} \right] = \left(\frac{\delta}{t'} \right)^2 + N^2 \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h}$$

so that

$$n = \frac{N^2 \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{w_h}}{\left(\frac{\delta}{t'} \right)^2 + N^2 \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h}} \qquad \text{-----(6.27)}$$

(6.27) give the optimum sample size in the case of a stratified random sample. To use the formula we require the sample stratum weight w_h. If the allocation of the total sample size n to the strata are made in proportional mode, we get $w_h = W_h$ and hence the formula can be used straight away. However, in the general case w_h's may have to be fixed in advance. Approximate choices of w_h can be made keeping in mind the desired representation to each stratum in the sample, based on the knowledge of the stratum totals and variances.

Fixing the stratum sample size

The formula (6.27) gives the optimum value of the overall sample size. Instead it is possible to fix each stratum size n_h independently and thus combine the values to get the total sample size. Treating each stratum independently, the optimum n_h can be computed using the formula (6.25) where N is replaced by N_h and S^2 by S_h^2, δ the tolerable margin of error in the estimate of the stratum total may be taken as same for all strata. Thus the optimum sample size is given by $n = \sum_{h=1}^{L} n_h$, where n_h is given by (6.25) as described above.

6.3.2 Optimum sample size for two stage PPS Design

Here we develop the procedure for determining the optimum size of the PSU and SSU in the two stage PPS random sampling design described in section 6.2 by using the criteria (b) indicated at the beginning of this section.

Under the two stage PPS design described in section 6.2, we have the unbiased estimator of the total landings per day given by (6.11) as

$$\hat{Y}_{ppz} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{T}\frac{M_{ik}\bar{y}_{ik.}}{p_i}$$

with variance given by (6.16)

$$V(\hat{Y}_{ppz}) = \frac{1}{n}\sum_{i=1}^{N}\sum_{k=1}^{T}\frac{M_{ik}^2}{p_i}\left(\frac{1}{m_{ik}}-\frac{1}{M_{ik}}\right)S_{ik}^2 + \frac{1}{n}\sum_{i=1}^{N}p_i\left(\frac{Y_{i.}}{p_i}-Y\right)^2$$

In order to estimate the optimum sizes of the PSU and SSU explicitly using criteria (b), we require the cost of the survey to be specified.

Let us assume that the total cost of collecting data from a zone can be regarded as made up of three components c_1, c_2 and c_3 where c_1 denote the fixed cost/PSU, c_2 the fixed cost per craft type considered and c_3 the cost/SSU.

Then the total cost for the survey as per the PPS design of section 6.2 will be

$$nc_1 + c_2\sum_{i=1}^{n}\sum_{k=1}^{T}M_{ik} + c_3\sum_{i=1}^{n}\sum_{k=1}^{T}m_{ik} \qquad \text{-----(6.28)}$$

The expected value of the total cost of the survey will be

$$C = nc_1 + nc_2\sum_{i=1}^{N}p_i\sum_{k=1}^{T}M_{ik} + nc_3\sum_{i=1}^{N}p_i\sum_{k=1}^{T}m_{ik} \qquad \text{-----(6.29)}$$

As per criteria (b) the sample size is determined so as to have $V(\hat{Y}_{ppz})$ subject to a fixed expected budget. We adopt a procedure similar to the one described in section 6.9 of Desraj (1971). In order to minimize $V(\hat{Y}_{ppz})$ subject to a fixed expected budget, we construct the function

$$G = V(\hat{Y}) + \lambda\left(nc_1 + nc_2\sum_{i=1}^{N}\sum_{k=1}^{T} p_i M_{ik} + nc_3\sum_{i=1}^{n}\sum_{k=1}^{T} p_i m_{ik} - C \right)$$

where λ is the Lagrangian multiplier.

Using (6.16),

$$G = \frac{1}{n}\sum_{i=1}^{N}\sum_{k=1}^{T_i}\frac{M_{ik}^2}{p_i}\left(\frac{1}{m_{ik}} - \frac{1}{M_{ik}}\right)S_{ik}^2 + \frac{1}{n}\sum_{i=1}^{N} p_i\left(\frac{Y_{i.}}{p_i} - Y\right)^2$$
$$+ \lambda\left(nc_1 + nc_2\sum_{i=1}^{N}\sum_{k=1}^{T_i} p_i M_{ik} + nc_3\sum_{i=1}^{n}\sum_{k=1}^{T_i} p_i m_{ik} - C \right) \qquad \text{-----(6.30)}$$

For G to be minimum, we must have, the partial derivatives of G with respect to m_{ik}, n and λ equal to zero.

$$\frac{\partial G}{\partial m_{ik}} = 0 \qquad \Rightarrow \frac{1}{n}\frac{M_{ik}^2}{p_i}\left(\frac{-1}{m_{ik}^2}\right)S_{ik}^2 + \lambda n c_3 p_i = 0$$

$$\Rightarrow \frac{M_{ik}^2 S_{ik}^2}{n\lambda c_3 p_i^2} = m_{ik}^2$$

$$\Rightarrow m_{ik} \propto \frac{M_{ik} S_{ik}}{p_i},$$

since n, λ and c_3 do not vary with respect to the subscripts i and k, they are treated as constants as far as the value of m_{ik} is concerned. Hence, we can take

$$m_{ik} = a\left(\frac{M_{ik} S_{ik}}{p_i}\right), \text{ for } i = 1,2,\ldots,n \;,\; k = 1,2,\ldots,T \qquad \text{-----(6.31)}$$

In order to simplify the solution of the estimating equations $\frac{\partial G}{\partial n} = 0$ and

$\frac{\partial G}{\partial \lambda} = 0$, we use the substitution

$$na = a' \qquad \text{-----(6.32)}$$

using (6.31) and (6.32) in (6.30), we get

$$G = \frac{1}{n}\left\{\sum_{i=1}^{N}\frac{1}{p_i}\sum_{k=1}^{T}\left(\frac{M_{ik}^2 S_{ik}^2}{aM_{ik}S_{ik}}\right)p_i - \sum_{i=1}^{N}\frac{1}{p_i}\sum_{k=}^{T}\left(\frac{M_{ik}^2 S_{ik}^2}{M_{ik}}\right)\right\} + \frac{1}{n}\sum_{i=1}^{N}\left(\frac{Y_{i.}}{p_i} - Y\right)^2$$
$$+ \lambda n\left(c_1 + c_2\sum_{i=1}^{N}\sum_{k=1}^{T}M_{ik}\right) + \lambda n c_3\left(\sum_{i=1}^{N}p_i\sum_{k=1}^{T}\frac{aM_{ik}S_{ik}}{p_i}\right) - \lambda C$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{N}\sum_{k=1}^{T}\left(\frac{M_{ik}S_{ik}}{a}\right) - \sum_{i=1}^{N}\frac{1}{p_i}\sum_{k=1}^{T}M_{ik}S_{ik}^2\right\} + \frac{1}{n}\sum_{i=1}^{N}p_i\left(\frac{Y_{i.}}{p_i} - Y\right)^2$$
$$+ \lambda n\left(c_1 + c_2\sum_{i=1}^{N}\sum_{k=1}^{T}M_{ik}\right) + \lambda n a c_3\left(\sum_{i=1}^{N}\sum_{k=1}^{T}M_{ik}S_{ik}\right) - \lambda C$$

$$= \frac{1}{a'}\sum_{i=1}^{N}\sum_{k=1}^{T}M_{ik}S_{ik} - \frac{1}{n}\left\{\sum_{i=1}^{N}\frac{1}{p_i}\sum_{k=1}^{T_i}M_{ik}S_{ik}^2 - \sum_{i=1}^{N}p_i\left(\frac{Y_{i.}}{p_i} - \hat{Y}\right)^2\right\} +$$
$$+ \lambda n\left(c_1 + c_2\sum_{i=1}^{N}p_i\sum_{k=1}^{T_i}M_{ik}\right) + \lambda a' c_3\left(\sum_{i=1}^{N}\sum_{k=1}^{T_i}M_{ik}S_{ik}\right) - \lambda C \qquad \text{-----(6.34)}$$

Treating (6.34) as a function containing the two unknowns n and a',we solve the equations $\frac{\partial G}{\partial n} = 0$ and $\frac{\partial G}{\partial a'} = 0$ to determine n and a'.

$$\frac{\partial G}{\partial n} = 0 \qquad \Rightarrow -\left(-\frac{1}{n^2}\right)\left\{\sum_{i=1}^{N}\frac{1}{p_i}\sum_{k=1}^{T}M_{ik}S_{ik}^2 - \sum_{i=1}^{N}p_i\left(\frac{Y_{i.}}{p_i} - Y\right)^2\right\}$$
$$+ \lambda\left(c_1 + c_2\sum_{i=1}^{N}p_i\sum_{k=1}^{T}M_{ik}\right) = 0$$

$$\frac{\partial G}{\partial n}=0 \qquad \Rightarrow \frac{1}{n^2}=\frac{\lambda\left(c_1+c_2\sum_{i=1}^{N}p_i\sum_{k=1}^{T}M_{ik}\right)}{\sum_{i=1}^{N}p_i\left(\frac{Y_{i.}}{p_i}-Y\right)^2-\sum_{i=1}^{N}\frac{1}{p_i}\sum_{k=}^{T}M_{ik}S_{ik}^2} \qquad \text{-----(6.34)}$$

$$\frac{\partial G}{\partial a'}=0 \qquad \Rightarrow\left(-\frac{1}{a'^2}\right)\sum_{i=1}^{N}\sum_{k=}^{T}M_{ik}S_{ik}+\lambda c_3\sum_{i=1}^{N}\sum_{k=1}^{T}M_{ik}S_{ik}=0$$

$$\Rightarrow a'^2=\frac{1}{\lambda c_3} \qquad \text{-----(6.35)}$$

Using (6.34) and (6.35), we get

$$\frac{a'^2}{n^2}=\frac{c_1+c_2\sum_{i=1}^{N}p_i\sum_{k=1}^{T}M_{ik}}{c_3\left(\sum_{i=1}^{N}p_i\left(\frac{Y_{i.}}{p_i}-Y\right)^2-\sum_{i=1}^{N}\frac{1}{p_i}\sum_{k=}^{T}M_{ik}S_{ik}^2\right)}$$

$$\Rightarrow a=\frac{\left(c_1+c_2\sum_{i=1}^{N}p_i\sum_{k=1}^{T}M_{ik}\right)^{\frac{1}{2}}}{\left(c_3\left(\sum_{i=1}^{N}p_i\left(\frac{Y_{i.}}{p_i}-Y\right)^2-\sum_{i=1}^{N}\frac{1}{p_i}\sum_{k=}^{T}M_{ik}S_{ik}^2\right)\right)^{\frac{1}{2}}} \qquad \text{-----(6.36)}$$

Using (6.36) in (6.31) we get

$$m_{ik}=\left(\frac{M_{ik}S_{ik}}{p_i}\right)\frac{\left(c_1+c_2\sum_{i=1}^{N}p_i\sum_{k=1}^{T}M_{ik}\right)^{\frac{1}{2}}}{\left(c_3\left(\sum_{i=1}^{N}p_i\left(\frac{Y_{i.}}{p_i}-Y\right)^2-\sum_{i=1}^{N}\frac{1}{p_i}\sum_{k=}^{T}M_{ik}S_{ik}^2\right)\right)^{\frac{1}{2}}} \qquad \text{----(6.37)}$$

for $i=1,2,\ldots,n$ and $k=1,2,\ldots,T$

Using (6.37) in (6.30), the value of n can be obtained as

$$n = \frac{C}{c_1 + c_2 \sum_{i=1}^{N} p_i \sum_{k=1}^{T} M_{ik} + c_3 \sum_{i=1}^{n} p_i \sum_{k=1}^{T} m_{ik}} \qquad \text{-----(6.38)}$$

This gives the cost components c_1, c_2, c_3 and the total budget C, the optimum sample size for PSU, n as well as that of SSU, m_{ik} can be obtained using the formula (6.37) and (6.38) respectively.

For using the above formulae, we require the total cost of the survey described by the function (6.28). It is true that at present the field staff and all others involved in the survey process are salaried employees and hence, the cost components c_1, c_2 and c_3 are not available explicitly. However, for the purpose of estimating the optimum sample size one has to split the total cost as described above.

6.4 Conclusion

The two stage PPS design developed in this chapter is operationally simple. Because of utilizing the past information for sample selection more precise estimator can be obtained even when the sampling fraction is slightly less than the targeted. This may also result in reduction in the total cost. The With Replacement scheme of selecting the PSU is used specifically for structural simplicity. For using the PPS design, we require the value of p_i for each centre. To get meaningful estimates corresponding to unobserved centres we require the count of the boats that are landed there. This can be obtained if there exists a mechanism for compulsory recording of all the boats operated at every centre.

The method developed for estimating the optimum sample size can be utilized for planning and administration of the survey. Though the optimum sample size is derived for the PPS design, given an appropriate cost function, they can be estimated in the case of other designs developed in chapter 5 in a similar manner. We are unable to include illustrations of the new design as well as the method of estimation for want of adequate existing data.

Concluding Remarks and Recommendations

The primary objective of this study was to critically evaluate the existing CMFRI sampling design and to suggest modifications to improve it. Rather than providing an estimator, we mainly concentrated on the scientific background of the methodology and reliability of the estimator. The study was mainly based on data collected through the regular surveys and periodic census by CMFRI. Due emphasis is given to utilize the schemes suitable to marine fishery sector from the available schemes of multi-stage sampling and estimation techniques.

The detailed study based on the available data revealed that

(i) The fishery data is highly unstable over time and hence does not yield to any scientific procedure of predictions even over short periods.

(ii) The total variability in the estimate of fish landings data in a zone is the joint effect of several components such as landing centre, landing day, fishing boat type and several other random factors.

(iii) The major share of the total variation is often due to the fishing boat type and several other random factors known as residual, which were not explicitly taken into account so far.

(iv) The sampling fraction attained is very low in all the zones.

The distinguishing feature of our study is that we could identify the major sources of variation in the highly unstable fishery data and propose new methods of estimation which accounts for these variations. A method for estimating the optimum sample size is also developed.

Based on the findings of the study, three new sampling designs are proposed – one intended for single centre zones and other two for multi-centre zones. The proposed first new design - two stage post-stratified design is more or less similar to the existing design, with an additional facility to account for the variation due to fishing unit types. This design though assumes post-stratification of the data based on gear types, stratification based on any other more suitable observable characteristics also can be considered equally well. The proposed second new design - three stage post-stratified design, though more scientific, may involve very high operational cost as every selected centre in a stratum must be observed for the same fixed number of days in a month. The significant feature of this design is that it ensures relatively higher sampling fraction in addition to accounting for variation due to all relevant sources. Subsequently, a third new design, which is operationally very simple and may cause a slight reduction in the sampling fraction by utilizing past information is proposed. Depending on the need and convenience, either the second or the third design can be chosen for multi-centre zones. The method developed for estimating the optimum sample size can be utilized for planning and administration of the survey. Though the optimum sample size are derived for the third design, given an appropriate cost function, they can be estimated in the case of other designs also in a similar manner.

On the basis of the findings of the study the following recommendations are made.

(i) Whatever is the sampling scheme adopted, the sampling fraction must be increased to the optimum level.

(ii) In the case of multi-centre zones, it is recommended to use the three stage post-stratified design. However, if structural simplicity and reduction of cost are of prime concern then the two stage PPS design may be adopted.

(iii) For all single centre zones the new two stage post-stratified design may be adopted.

(iv) There has to be a compulsory recording system at all landing centres for all fishing trips whenever they go for fishing. The record may contain details on the category of the fishing unit, capacity, manpower, total catch, craft details, actual fishing hours, landing time etc. This will provide the actual information on the count and category of the fishing units which form a part of the estimator. This will increase the accuracy of the estimator and also cause reduction in its variance.

(v) In the case of getting actual data on the number of boats operated at each centre, more reliable estimators can be developed even when the sampling fraction is very small.

(vi) The present practice of splitting an observation day into three periods may be discontinued. Instead the observation day may be considered as a single unit. The count of fishing units may be recorded continuously during the day of observation, while recording the landings is to be restricted to the forenoon and afternoon sessions only.

(vii) The selection of fishing units for recording landings may be made at intervals of 15 to 20 minutes during the time of field visit with priority for distinct gear types if available. In the case of no new gear types available, priority is to be given to get at least two or three boats of the same gear. The night landings are to be considered only for recording count by enquiry in the forenoon of the following day of visit. Treating the 24 hour duration of a day into a single unit will give more freedom to the filed staff to ensure adequate representation to each distinct gear type operated on the day. The forenoon session of the following day can be mainly targeted to select new gear types as well as to give adequate representation to the already noted gear types.

Details of the Papers Presented and Conference/Workshop Attended

As part of this study, we have made the following presentations at national/international seminars.

1. Presented the paper titled **Design of a system for the collection of marine fish landings in the state of Kerala, India** by Mini K.G. and M. Kumaran at the *International Indian Statistical Association (IISA) Joint Statistical Meeting and International Conference on Statistics, Probability and Related Areas* organized by Department of Statistics, Cochin University of Science and Technology, Cochin during January 2-5, 2007.

   Presented the paper titled **Estimation of Variance in The Marine Fish Landings Data By Nested Sampling Method** by Mini K.G. and M. Kumaran in the *National Seminar on Recent Advances in Statistics & Analysis of Non-conventional data* organized by Department of Statistics, Farook College, Kozhikode and Kerala Statistical Association during March 15-17, 2008.

   Presented the paper titled **On Estimation and Testing of Variance Components in Hierarchical Data** by M. Kumaran and Mini K.G. and in the *National Seminar on Recent Advances in Statistics & Analysis of Non-conventional data* organized by Department of Statistics, Farook College, Kozhikode and Kerala Statistical Association during March 15-17, 2008.

Publications

The following articles are communicated for publication in international journals.

1. **Estimation of Variance Components in the Marine Fish Landings Data by the Method of Nested Analysis** submitted for publication to *Canadian Journal of Fisheries and Aquatic Science.*

2. **On Use of Post-stratification for Estimating the Marine Fish Landings** submitted for publication to *Indian Journal of Marine Science.*

3. **A Three Stage Post-stratified Design for Estimation of Marine Fish Landings** submitted for publication to *Fisheries Research.*

4. **A Two stage PPS Design for Estimating the Marine Fish Landings** submitted for publication to *Biometrics.*

REFERENCES

Akar, I. and Sedransk, J. 1979. Post-stratified cluster sampling. *Sankhya*, C. 41(1): 76-83.

Alagaraja, K. 1984. Simple methods for estimates of parameter for assessing exploited fish stocks. *Indian J. Fish.*, 31(2): 177-208.

Ammini, P.L. 1999. Status of marine fisheries in Kerala with reference to ban of monsoon trawling. *Marine Fisheries Information Services T & E Ser.* 160: 25-36.

Anderson, R. L. and Bancroft, T.A. 1952. *Statistical Theory in Research,* McGraw-Hill, New York.

Anderson R. L. 1965. Query Number 9. *Technometrics,* 7: 75-76.

Anscombe, F. J. 1948. The transformation of Poisson, binomial and negative-binomial data. *Biometrika,* 35: 246–254.

Bal, D.V. and Banerji, S.K. 1951. A survey of the sea fisheries of India. *Proc. Indo-Pacific Fish Council.* Sec. II: 176-181.

Balan, K., Panikkar, K. K. P., Jacob, T. Andrews, J. and Rajendran, V. 1989. *Motorization of country craft in Kerala- An Impact Study.* CMFRI Spl. Publ., 45: 74 pp.

Banerjii, S. K. and Chakraborty, D. 1972. On estimation of yield from exploited marine stocks with reference to South East Asia. In : *Proceedings of symposium on Living Resources of the Seas around India,* CMFRI., Spl. Publ., 176-183.

Bennett, C. A. and Franklin, N. L. 1954. *Statistical Analysis in Chemistry and the Chemical Industry.* Wiley, New York.

Brander, K. 1975. Guidelines for collection and compilation of fishery statistics. *FAO. Fish. Tech. Pap.* 148: 46pp.

CMFRI. 1981. All India Census of Marine Fishermen, Craft and Gear: 1980. *Marine Fisheries Information Services T & E Ser.* 30: 2-30.

CMFRI. 1983. Proceedings of the Workshop on Acquisition and Dissemination of Data on Marine Living Resources of Indian Seas. *Marine Fisheries Information Services T & E Ser.* 46: 1-49.

Cochran, W. G, 1946. Relative accuracy of systematic and stratified random sampling for a certain class of populations. *Ann. Math. Stat.*, 17: 164-177.

Cochran, W. G. 1977. *Sampling Techniques.* Third Edition, John Wiley & Sons, New York.

CSO. 1961. *National Income Statistics.* Central Statistical Organisation, Government of India.

Das, A. C. 1950. Two dimensional systematic sampling and associated stratified and random sampling. *Sankhya,* 10: 95-108.

DAHD&F, 2006. *Marine Fisheries Census, 2005* - Kerala. Department of Animal Husbandry, Dairying and Fisheries, Ministry of Agriculture, Government of India and Central Marine Fisheries Research Institute, Kochi. Part-III(6), 217 pp.

D'Cruz, S. T. 1998. *The Ring Seine: Evolution and Design Specifications.* SIFFS Publication, Thiruvananthapuram.

Desraj, D. 1971. *Sampling theory.* McGraw-Hill, New York.

Durbin, J. 1967. Design of multi-state surveys for the estimation of sampling errors. *Applied Statistics,* 16: 52-164.

Federerer, W. T., 1968 Non-Negative Estimators for Components of Variance, *Applied Statistics,* 17(2): 171-174.

Fletcher, D. J. and Underwood, A. J. 2002. How to cope with negative estimates of components of variance in ecological field studies. *J. Exp. Mar. Biol. Ecol.,* 273: 89-95.

Fuller, W.A. 1966. Estimation employing post strata, *J. Amer. Stat. Assoc.,* 61: 1172-1183.

Gagnon, F., Lee, H., Rancourt, E. and Särndal, C. E. 1997. Estimating the variance of the generalised regression estimator in the presence of imputation for the generalised estimation system. *Proceedings of the survey methods section, Statistical Society of Canada,* 151–156.

Ganguli, M. 1941. A note on nested sampling. *Sankhya,* 5: 449-452.

Gates, C. E. and Shiue, C. 1962. The analysis of variance of the S-stage hierarchal classification. *Biometrics,* 18: 529-536.

Gaylor, D. W. and Hartwell, T. D. 1969. Expected mean square for nested classifications. *Biometrics,* 25: 427-430

Government of India 1951. *Report on the Marketing of Fish in the Indian Union* - Directorate of Marketing and Inspection, Government of India.

Gower, J. C. 1962. Variance component estimation for unbalanced hierarchal classifications. *Biometrics,* 18: 529-536.

Gulland, J. G. 1955. *Estimation of growth and mortality in commercial fish populations*. Ministry of Agriculture and Fisheries, Fishery Investigations, Series II, XVII, 46pp.

Hansen, M. H. and Hurwitz, W. N. 1943. On the theory of sampling from finite populations. *Ann. Maths. Stat.*, 14: 333-362.

Hansen, M. H. and Hurwitz, W. N. 1949. On the determination of optimum probabilities in sampling. *Ann. Maths. Stat.*, 20: 426-432.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. 1953. *Sample Survey Methods and Theory.* Volumes I & II, John Wiley & Sons, New York.

Herbach, L. H. 1959. Properties of model II-type analysis of variance tests. *Ann. Math. Statist.* 30: 939–959.

Holt, D. and Smith, T. M. F. 1979. Post-stratification. *J. Roy. Stat. Soc. A*, 142(1): 33-46.

ICAR.1965. *Sample surveys for Improvement of Marine Fishery Statistics.* Technical bulletin (Anim.Husb.). Indian Council of Agricultural Research, New Delhi, 2: 87pp.

Jacob, T., Rajendran, V., Pillai, P. K. M., Andrews, J. and Satyavan, U. K. 1983. *An appraisal of the marine fisheries in Kerala.* CMFRI Spl. Publ., 35: 42pp.

Khuri, A. I. 2000. Designs for Variance Components Estimation: Past and Present . *Int. Stati. Rev.,* 68 (3): 311–322.

Krishnaiah, P. R. and Rao, C. R. Editors. 1988. *Handbook of Statistics,* Vol. 6, Sampling. Elsevier Science Publications, Amsterdam.

Kumar, R. 1981. Some investigations in two dimensional sampling. *Unpublished M.Sc. (Ag. Stat.) thesis,* IARI, New Delhi.

Kumar, K. 1989. On post-stratification for improvement of estimation procedure in a multistage sampling design. *Unpublished Ph.D thesis,* IARI, New Delhi.

Kurien, J and Willmann R. 1982. *The Costs and Earnings of Artisanal and Mechanized Fisheries of Kerala State.* FAO/UNPD/BOBP Publication, Madras.

Kurien, J. 1985. Technical Assistance Projects and Socio-Economic Change: The Norwegian Intervention in Kerala's Fisheries Development Experience. *Economic and Political Weekly,* 20: 25-26.

Kurien, J. 2000. *Factoring Social and Cultural Dimensions into Food and Livelihood Security Issues of Marine Fisheries, A Case Study of Kerala State, India,* Centre For Development Studies, Working Paper No. 299, February, 50pp.

Kutty, M. K., Nair, A. K. K., and Qasim, S. Z. 1973. An evaluation of the sampling design adopted by the CMFRI for estimating marine fish production in India. *Indian J. Fish.* 22: 16-34.

Mahajan, R. K. 1984. Sampling on successive occasions in two dimensional populations. *Unpublished Ph.D thesis,* IARI, New Delhi.

Mahalanobis, P. C. 1946. Recent experiments in statistical sampling in the Indian Statistical Institute, *J. Roy. Stat. Soc. (A),* 109: 325-370.

Mahalanobis, P. C. 1952. Some aspects of the design of sample surveys, *Sankhya,* 12: 1-7.

Mahamunulu, D. M. 1963. Sampling variances of the estimates of variance componets in the unbalanced 3-way nested classification. *Ann. Math. Statist.*, 34: 521-527.

McHugh, R. B. and Mielke, P. W. 1968. Negative variance estimates and statistical dependence in nested sampling. *J. Amer. Stat. Assoc.*, 63: 1000-1003.

Mehrotra, P. C., Srivastava, A. K. and Tyagi, K. K. 1984. On post-stratification for cluster sampling. *J. Ind. Soc. Agril. Stat.*, 36(3): 98-104.

Mehrotra, P. C. 1993. A scheme in post stratification in two stage sampling. *J. Ind. Soc. Agril. Stat.*, 45: 163-70.

Miller, T. J., Skalski, J. R., and Ianelli, J. N. 2007. Optimizing a stratified sampling design when faced with multiple objectives. *ICES J. Mar. Sci.*, 64(1): 97–109.

Murthy, M.N. 1967. *Sampling Theory and Methods.* Statistical Publishing Society, Calcutta, India.

Narang, M. S. 1994. Some studies in post-stratified design involving auxiliary information. *Unpublished Ph.D thesis*, IARI, New Delhi.

Nelder, J. A. 1954. The interpretation of negative components of variance. *Biometrika*, 41: 544-548.

Panse, V. G. and Sastry, K. V. R. 1960. *Sample surveys for fishery statistics.* Technical Report to the Government of United Arab Republic, ETAP, FAO, 1247: 96pp.

Papaconstantinou, C., Kavadas, S., Tsimenidis, N., Economou, A. N. and Bazigos, G. 2002. Design of a system for the collection and compilation of basic fisheries statistics in the Mediterranean: a case study for Greece. *Fisheries Research,* 57: 155-167.

Pfeffermann, D. and Krieger, A. M. 1991. Poststratification using regression estimators when information on strata means and sizes is missing. *Biometrika,* Vol. 78 (2): 409-419.

Quenoullie, M. H. 1949. Problems in plane sampling. *Ann. Math. Stat.,*20: 355-75.

Särndal, C. E., Swensson, B and Wretman, J. 1992. *Model Assisted Survey Sampling.* Springer-Verlag, New York, 694pp.

Sen, A. R. 1986. Methodological problems in sampling commercial rockfish landings. *Fishery Bulletin,* 84: 409-421.

Sen, A. R. 1990. Sampling methodology for a boat-fishing survey design for Hawaii. *Biom. J.* 2: 155-176.

Sethi, A. S. and Srivastava, A. K. 1987. Ratio estimator with post-stratification design involving double sampling approach. *J. Ind. Soc. Agril. Stat.* 39: 34-48.

SIFFS. 1992. *A Census of the Artisanal Marine Fishing Fleet of Kerala State.* SIFFS Publications, Thiruvananthapuram.

Singh, D. 1954. The sampling with varying probabilities without replacement. *J. Ind. Soc. Agril. Stat.,* 6: 48-57.

Singh, M. and Gupta, V. K. 1980. On the estimation of mean parameter with random number of observations on each unit. *Biom. J.*, 22(1): 41-45.

Southward, G. M. 1976. Sampling landings of halibut for age composition. *Int. Pac. Halibut. Comm. Sci. Rep.*, 58: 1-31.

Srinath, M., Somy Kuriakose and Mini K. G., 2005. *Methodology for estimation of marine fish landings in India.* CMFRI Spl. Publ., No. 86: 57pp.

Srinath, M. and Jayasankar, J. 2007. Advancement in research on fish stock assessment. In: *Status and Perspectives in Marine Fisheries Research in India,* Mohan Joseph Modayil and N.G.K. Pillai (Eds.) Central Marine Fisheries Research Institute, Kochi, 196-210.

Stamatopoulos, C. 2002. Sample based fishery surveys: A technical handbook. *FAO. Fish. Tech. Pap.* 425: 132pp.

Sukhatme, P. V. 1947. The problem of plot size in large scale yield surveys. *J. Amer. Stat. Assoc.,* 42: 297-310.

Sukhatme, P. V. 1950. Sample surveys in agriculture, Presidential Address to the Statistical Section. *Proceedings of the Indian Science Congress*, 37th Session, Pune.

Sukhatme, P. V. and Panse, V. G. 1951. Crop surveys in India II, *J. Ind. Soc. Agril. Stat.*, 3: 97-168.

Sukhatme, P. V. and Narain, R. D. 1952. Sampling with replacement. *J. Ind. Soc. Agril. Stat.*, 4: 42-49.

Sukhatme, P. V., Panse, V.G. and Sastry, K. V. R. 1958. Sampling techniques for estimating catch of sea fish in India. *Biometrics*, 14: 78-96.

Sukhatme, P. V. and Sukhatme, B. V. 1976. *Sampling Theory of Surveys with Applications. J. Ind. Soc. Agril. Stat.*, New Delhi.

Sukhatme, P. V, Sukhatme, B. V, Sukhatme, S. and Asok, C. 1997. *Sampling Theory of Surveys with Applications. J. Ind. Soc. Agril. Stat.*, New Delhi.

Thompson, S. K. 2000. *Sampling*. John Wiley & Sons, New York.

Thompson, W.A. Jr., 1962. The problem of Negative estimates of variance components. *Ann. Math. Stat.*, 33: 273-289.

Thompson, W. A. and Moore, J. R. 1963. Non negative estimates of variance components. *Technometrics*, 5(4): 441-449.

Tomlinson, P. K. 1971. Some sampling problems in fishery work. *Biometrics,* 27: 631-641.

Vijayan, V., Edwin, L. and Ravindran, K. 2000. Conservation and management of Marine Fishery Resources of Kerala State, India. *Naga*, The ICLARM Quarterly, 23(3): 6-9.

Yates, F. 1949. *Sampling Methods for Censuses and Surveys*, Charles Griffin and Co., London.

Williams, W. H. 1962. The variance of an estimator with post stratified weighting. *J. Amer. Stat. Assoc.* 57: 622-27.