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 \( m_{n(j)} \) 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 \( M_{n(j)} \), 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
where \( y_{a(j)} \) 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) \( y_{a(j)} \) 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

\[
\hat{\gamma}_{pos} = \sum_{j=1}^{T} N_{(j)} \frac{n_{(j)}}{n} \sum_{i=1}^{m_{(j)}} \sum_{k=1}^{m_{(j)}} y_{a(j)}, \quad \text{(5.2)}
\]

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

\[
\text{var}(\hat{\gamma}_{pos}) = \sum_{j=1}^{T} N_{(j)}^2 \left( \frac{1}{n} - \frac{1}{N} \right) w_{(j)}^2 s_{(j)}^2 + \left[ \frac{N_{(j)}^2}{n^2} \sum_{j=1}^{T} (1 - w_{(j)}) s_{(j)}^2 \right]
\]

\[
+ N^2 \sum_{j=1}^{T} w_{(j)} \left( \left( 1 + \frac{1 - w_{(j)}}{w_{(j)}} + \frac{1 - w_{(j)}}{w_{(j)}} + \frac{1 - w_{(j)}}{w_{(j)}} \right) - \left( 1 + \frac{w_{(j)}}{N_{(j)}} \right) \right) s_{(j)}
\]

\[
+ N^2 \sum_{j=1}^{T} \frac{1}{nN} \left( 1 + \frac{1 - w_{(j)}}{w_{(j)}} \right) \sum_{i=1}^{m_{(j)}} \left( \frac{M_i}{w_{(j)}} \left( \frac{1}{m_i} - \frac{1}{M_i} \right) + \frac{1 - w_{(j)}}{m_i^2} \right) s_{(j)}^2 \quad (5.3)
\]

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

\[
w_{(j)} = \frac{N_{(j)}}{N}, \quad w_{(j')} = \frac{N_{(j')}}{N}, \quad w_{(y')} = \frac{N_{(y')}}{N}, \quad w_{(y)} = \frac{M_{(y)}}{M_i}
\]
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 \( V(\hat{\lambda}_{\text{part}}) \) is given by

\[
\hat{V}(\hat{\lambda}_{\text{part}}) = \sum_{j=1}^{L} \frac{N_j}{n} \left( \frac{1}{n} - \frac{1}{N} \right) w_{ij} S_{i(j)}^2
\]

\[+ \sum_{j \neq j'} N_{(j)} N_{(j')} \left( \frac{n_{(j)} n_{(j')}}{n_{(j)} n_{(j')}} - \frac{n_{(j')}}{n_{(j')}} \frac{n_{(j)}}{n_{(j)}} \right) S_{i(j')}
\]

\[+ \sum_{j=1}^{L} \sum_{i=1}^{n_{(j)}} M_{(j)} \left( \frac{1}{m_{(j)}} - \frac{1}{M_{(j)}} \right) S_{i(j)}^2, \] \quad (5.7)

where

\[
S_{i(j)}^2 = \frac{1}{n_{(j)} - 1} \sum_{i=1}^{n_{(j)}} (y_{i(j)} - \bar{y}_{i(j)})^2
\] \quad (5.8)
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>
<thead>
<tr>
<th>Month</th>
<th>Existing design</th>
<th>Design I</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Variance</td>
</tr>
<tr>
<td>Jan</td>
<td>362</td>
<td>4112</td>
</tr>
<tr>
<td>Feb</td>
<td>479</td>
<td>11301</td>
</tr>
<tr>
<td>Mar</td>
<td>883</td>
<td>100427</td>
</tr>
<tr>
<td>Apr</td>
<td>531</td>
<td>43730</td>
</tr>
<tr>
<td>Jul</td>
<td>551</td>
<td>63599</td>
</tr>
<tr>
<td>Aug</td>
<td>4895</td>
<td>120478</td>
</tr>
<tr>
<td>Sep</td>
<td>3460</td>
<td>1754069</td>
</tr>
<tr>
<td>Oct</td>
<td>2347</td>
<td>235540</td>
</tr>
<tr>
<td>Nov</td>
<td>631</td>
<td>47527</td>
</tr>
<tr>
<td>Dec</td>
<td>457</td>
<td>3546</td>
</tr>
</tbody>
</table>

* 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 1, 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 $m_{hkij}$ and $M_{hkij}$ 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,

\[ m_{hkij} = \sum_{k=1}^{T} m_{hkij} \text{ and } M_{hkij} = \sum_{k=1}^{T} M_{hkij} \]

\[ m_{hij} = \sum_{j=1}^{D} m_{hij} \text{ and } M_{hij} = \sum_{j=1}^{D} M_{hij} \]

\[ m_h = \sum_{i=1}^{N} m_{hij} \text{ and } M_h = \sum_{i=1}^{N} M_{hij} \]

\[ m = \sum_{h=1}^{L} m_h \text{ and } M = \sum_{h=1}^{L} M_h \]
Let

\[ Y_{hijk} \] be the total quantity of fish landed by the \( t^{th} \) 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 \( Y_{hijk} \) is the species-wise sum of the observations on the day)

\[ Y_{hijk} = \sum_{j=1}^{J_{ijk}} Y_{hijk} \]

\[ Y_{hi} = \sum_{k=1}^{K} Y_{hijk} \]

\[ Y_{hi} = \sum_{j=1}^{J_{h}} Y_{hi} \]

\[ Y_{h} = \sum_{i=1}^{I_{h}} Y_{hi} \]

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

\[ \hat{Y}_M = \frac{L}{\sum_{h=1}^{n_h} d_h} \sum_{h=1}^{n_h} d_h \sum_{j=1}^{J_{hi}} \sum_{k=1}^{K} M_{hijk} \sum_{l=1}^{M_{hijk}} y_{hijkl} \]

The gear estimator is given by

\[ \hat{Y}_{MK} = \frac{L}{\sum_{h=1}^{n_h} d_h} \sum_{h=1}^{n_h} d_h \sum_{j=1}^{J_{hi}} \sum_{k=1}^{K} M_{hijk} \sum_{l=1}^{M_{hijk}} y_{hijkl} \]

Then

\[ E(\hat{Y}_M) = E_i E_j E_k (\hat{Y}_M) \]
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,

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

$$= E_1 E_2 \left( \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i=1}^{d_h} \sum_{j=1}^{d_h} \sum_{k=1}^{r} \frac{M_{hijk}}{m_{hijk}} \sum_{l=1}^{m_{hijk}} y_{hijkl} \right)$$

$$-----(5.12)$$

Now

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

$$= M_{hij} \sum_{k=1}^{r} W_{hijk} \bar{y}_{hijk}$$

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

$$= y_{hij}$$

$$-----(5.13)$$

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

145
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

\[
E(\hat{Y}_M) = E_i \left( \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{a_h} \frac{D_i}{d_h} \sum_{j=1}^{D_i} E_1 (Y_{ijn}) \right)
\]

\[
E(\hat{Y}_M) = E_i \left( \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{a_h} \frac{D_i}{d_h} \sum_{j=1}^{D_i} \left( \frac{1}{D_i} \sum_{j=1}^{D_i} Y_{ijn} \right) \right)
\]

\[
= \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{a_h} E_1 (Y_{ih})
\]

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

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

\[= Y \quad -----(5.14)\]
\[ V(\hat{Y}_M) = V_1 E_2 E_3 (\hat{Y}_M) + E_1 V_2 E_3 (\hat{Y}_M) + E_1 E_2 V_3 (\hat{Y}_M), \quad \text{(5.15)} \]

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),

\[
V_1 E_2 E_3 (\hat{Y}_M) = V_1 E_2 E_3 \left( \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} D_i \sum_{j=1}^{d_i} \sum_{k=1}^{r} \frac{M_{hijk}}{m_{hijk}} \sum_{l=1}^{m_{hijk}} Y_{hijkl} \right)
\]

\[
= V_1 \left( \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} D_i \sum_{j=1}^{d_i} \sum_{k=1}^{r} E_2 \left( \frac{T}{m_{hijk}} E_3 \left( \sum_{l=1}^{m_{hijk}} Y_{hijkl} \right) \right) \right) \quad \text{by (5.13)}
\]

\[
= V_1 \left( \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} D_i E_2 \left( \sum_{j=1}^{d_i} Y_{hij} \right) \right)
\]

\[
= V_1 \left( \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} Y_{hi} \right)
\]

\[
= \sum_{h=1}^{H} \frac{N_h}{n_h} \left( \frac{1}{n_h} - \frac{1}{N_h} \right) S_h^2,
\quad \text{(5.16)}
\]

where

\[
S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{n_h} (Y_{hi} - \bar{Y}_h)^2,
\]

since the first stage units selection is according to SRSWOR.

Now the second term of (5.15) is
\[ E_1V_2E_3(\bar{Y}_M) = E_1V_2E_3 \left( \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{l=1}^{n_h} \frac{D_h}{d_h} \sum_{j=1}^{d_h} \sum_{k=1}^{d_h} \frac{M_{hijk}}{m_{hijk}} \sum_{i=1}^{y_{hijkl}} \right) \]

\[ E_1V_2E_3(\bar{Y}_M) = E_1V_2 \left( \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{l=1}^{n_h} \frac{D_h}{d_h} \sum_{j=1}^{d_h} \sum_{k=1}^{d_h} \frac{Y_{hij}}{Y_{hijkl}} \right) \]

\[ = E_1 \left( \sum_{h=1}^{L} \frac{N_h^2}{n_h^2} \sum_{l=1}^{n_h} V_2 \left( \frac{D_h}{d_h} \sum_{j=1}^{d_h} \sum_{k=1}^{d_h} \frac{Y_{hij}}{Y_{hijkl}} \right) \right) \]

\[ = E_1 \left( \sum_{h=1}^{L} \frac{N_h^2}{n_h^2} \sum_{l=1}^{n_h} D_h^2 V_2(\bar{Y}_h) \right) \]

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

where

\[ S_{hi}^2 = \frac{1}{D_h - 1} \sum_{i=1}^{d_h} (Y_{hij} - \bar{Y}_h)^2 \]  

\[ \text{-----(5.17)} \]

since the second stage selection also is as per SRSWOR.

Hence,

\[ E_1V_2E_3(\bar{Y}_M) = \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{l=1}^{n_h} D_h^2 \left( \frac{1}{d_h} - \frac{1}{D_h} \right) S_{hi}^2 \]  

\[ \text{-----(5.18)} \]

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

Let \( \bar{y}_h \) 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 \( \bar{y}_{\mu g} = \sum_{g=1}^{L} W_g \bar{y}_g \) is the post-stratified estimator of the population mean. Then

\[
V(\bar{y}_{\mu g}^*) = EV(\bar{y}_{\mu g}^*) + V(\bar{y}_{\mu g}^*)/n_g
\]

\[
= EV(\bar{y}_{\mu g}^*)/n_g,
\]

since \( E(\bar{y}_{\mu g}^*/n_g) \) is a constant.

\[
V(\bar{y}_{\mu g}^*) = E\left( \sum_{g=1}^{L} W^2_g \left( \frac{1}{n_g} - \frac{1}{N_g} \right) S_g^2 \right)
\]

\[
= \sum_{g=1}^{L} \left( E\left( \frac{1}{n_g} \right) - \frac{1}{N_g} \right) W^2_g S_g^2 \quad \text{------(5.19)}
\]

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

\[
E\left( \frac{1}{n_g} \right) = \frac{1}{n W_g} + \frac{1 - W_g}{n^2 W_g^2}
\]

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

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

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

We get

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

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

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

by neglecting terms in higher powers of \( \Delta_g \) than the second.

Also \( E(\Delta_g) = 0 \)

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

Thus we get

\[ E\left( \frac{1}{n_g} \right) = \frac{1}{E(n_g)} \left( 1 + \frac{V(n_g)}{\left[ E(n_g) \right]^2} \right) \]

\[ \text{-----(5.20)} \]

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

\[ V(\bar{y}_{\mu}) = \sum_{g=1}^{L} \left( 1 + \frac{V(n_g)}{\left[ E(n_g) \right]^2} \right) \frac{1}{N_g} W_g^2 S_g^2 \]

\[ \text{------(5.21)} \]

In our present case \( n_g = m_{hij}, N_g = M_{hij}, \ n = m_{hij} \) and \( N = M_{hij} \) so that

\[ E(m_{hij}) = m_{hij} \frac{M_{hij}}{M_{hij}} \]

\[ \text{------(5.22)} \]
Now the third term of (5.15) is

\[ E_i E_2 V_3(\hat{Y}_{hl}) = E_i E_2 V_3 \left( \sum_{k=1}^{l_n} \frac{N_k}{n_k} \sum_{i=1}^{n_k} D_{hi} \sum_{j=1}^{d_k} \sum_{k=1}^{\frac{t_k}{k}} M_{hijk} m_{hijk} \sum_{l=1}^{r_{hijk}} y_{hijkl} \right) \]

\[ = E_i E_2 \sum_{k=1}^{l_n} \frac{N_k}{n_k} \sum_{i=1}^{n_k} \frac{D_k^2}{d_k^2} \sum_{j=1}^{d_k} V_{k} \left( M_{hijk} \sum_{k=1}^{t_k} W_{hijk} \hat{y}_{hijk} \right) \]

\[ = E_i E_2 \sum_{k=1}^{l_n} \frac{N_k^2}{n_k^2} \sum_{i=1}^{n_k^2} \frac{D_k^2}{d_k^2} \sum_{j=1}^{d_k} M_{hijk}^2 V_3(\hat{y}_{hijk}) \]

Treating \( \hat{y}_{hijk} \) as \( y_{hijk}' \), its variance is given by equation (5.21). Thus we get the above as

\[ E_i E_2 V_3(\hat{Y}_{hl}) = E_i E_2 \left( \sum_{k=1}^{l_n} \frac{N_k^2}{n_k^2} \sum_{i=1}^{n_k^2} \frac{D_k^2}{d_k^2} \sum_{j=1}^{d_k} M_{hijk}^2 \sum_{k=1}^{t_k} \frac{M_{hijk}}{m_{hijk} M_{hijk}} \right) \]

\[ \left[ 1 + \left( \frac{m_{hijk} M_{hijk}}{M_{hijk}} \right)^2 \right] \left[ \frac{1}{1 - \left( \frac{m_{hijk} M_{hijk}}{M_{hijk}} \right)^2} W_{hijk}^2 S_{hijk}^2 \right] \]

\[ = E_i E_2 \left( \sum_{k=1}^{l_n} \frac{N_k^2}{n_k^2} \sum_{i=1}^{n_k^2} \frac{D_k^2}{d_k^2} \sum_{j=1}^{d_k} \left( \frac{1}{M_{hijk}} - \frac{1}{M_{hijk}} \right) \sum_{k=1}^{t_k} M_{hijk}^2 S_{hijk}^2 \right) + \frac{M_{hijk}}{m_{hijk}} \left( \frac{M_{hijk} - m_{hijk}}{M_{hijk} - 1} \sum_{k=1}^{t_k} \left( 1 - \frac{M_{hijk}}{M_{hijk}} \right) S_{hijk}^2 \right) \]
where

$$S^2_{hijk} = \frac{1}{M_{hijk} - 1} \sum_{l=1}^{M_{hijk}} (Y_{hijk} - \bar{Y}_{hijk})^2$$

----- (5.24)

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

----- (5.25)

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

$$V(\hat{Y}_m) = \sum_{h=1}^{L} \left\{ N_h^2 \left( \frac{1}{n_h} - \frac{1}{N_h} \right) S^2_h + \frac{N_h}{n_h} \sum_{i=1}^{n_h} D_h \left( \frac{1}{d_h} - \frac{1}{D_h} \right) S^2_{hi} \right\} + \frac{N_h}{n_h} \sum_{i=1}^{n_h} D_h \left[ \left( \frac{1}{m_{hij}} - \frac{1}{M_{hij}} \right) \sum_{k=1}^{T} M_{hijk} S^2_{hijk} \right] + \frac{M_{hij}}{m_{hij}} \left( \frac{M_{hij} - m_{hij}}{M_{hij} - 1} \right) \sum_{k=1}^{T} \left( 1 - \frac{M_{hijk}}{M_{hij}} \right) S^2_{hijk}$$

----- (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 $S^2_h$, $S^2_{hi}$ and $S^2_{hijk}$ with their corresponding sample estimators $s^2_h$, $s^2_{hi}$ and $s^2_{hijk}$.
\[ \hat{V}(\hat{Y}_M) = \sum_{h=1}^{L} \left[ N_h^2 \left( \frac{1}{n_h} - \frac{1}{N_h} \right) s_h^2 + \frac{N_h}{n_h} \sum_{i=1}^{n_h} D_h \left( \frac{1}{d_h} - \frac{1}{D_h} \right) s_{hij}^2 \right] + \frac{N_h}{n_h} \sum_{i=1}^{n_h} D_h \sum_{j=1}^{D_h} \left[ \frac{1}{M_{hij}} - \frac{1}{m_{hij}} \right] \sum_{k=1}^{T} M_{hijk} s_{hijk}^2 + \frac{M_{hij}}{m_{hij}} \left( \frac{M_{hij} - m_{hij}}{M_{hij} - 1} \right) \sum_{k=1}^{T} \left( 1 - \frac{M_{hijk}}{M_{hij}} \right) s_{hijk}^2 \right] \] 

\[ - (5.27) \]

where

\[ s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left( y_{hi} - \bar{Y}_h \right)^2 \] 

\[ s_{hij}^2 = \frac{1}{d_h - 1} \sum_{i=1}^{d_h} \left( y_{hij} - \bar{Y}_{hi} \right)^2 \] 

\[ s_{hijk}^2 = \frac{1}{m_{hijk} - 1} \sum_{i=1}^{m_{hijk}} \left( y_{hijkl} - \bar{Y}_{hijk} \right)^2 \] 

\[ - (5.28) \]

\[ - (5.29) \]

\[ - (5.30) \]

**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 \( \hat{Y}_s \) 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

\[ V(\hat{Y}_Z) = \sum_{z=1}^{Z} \hat{V}(\hat{Y}_z) + \sum_{z=1}^{Z} \hat{V}(\hat{Y}_{Mz}) \]

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 \( f' = \frac{\sum n_h d_h}{\sum h=1 \frac{1}{N_h D_h}} \). The sampling fraction that can be achieved in an equivalent existing design is \( f = \frac{\sum n_h}{\sum h=1 \frac{1}{N_h D_h}} \). Suppose we fix \( d_h = c \) for all \( h \). Then \( f' = f \cdot c \). 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.