8.0. **Summary.**

This chapter considers the problem of optimum allocation of sample size to the strata when the finite population model furnished by Avadhani and Sukhatme (1970) holds.

8.1. **Introduction.**

One of the designs frequently used in sample surveys is stratified random sampling. Consider the population under study to be divided into $k$ strata each of size $N_i$, $i=1, 2, \ldots, k$ and a stratified random sample of size $n_i$, $i=1, 2, \ldots, k$ is drawn without replacement from the $i$th stratum so that $\sum_{i=1}^{k} n_i = n$. Let $\bar{Y}$ denote the characteristic of interest and $\bar{X}$ a characteristic which is highly correlated with $\bar{Y}$. Let $Y_{ij}$ and $X_{ij}$ be the $\bar{Y}$ and $\bar{X}$ characteristic values respectively of the $j$th unit in the $i$th stratum ($j=1, 2, \ldots, N_i$, $i=1, 2, \ldots, k$).

The population mean $\bar{Y}_N$ can be expressed as

$$
\bar{Y}_N = \frac{1}{N} \sum_{i=1}^{k} N_i \frac{1}{n_i} \sum_{i=1}^{N_i} Y_{ij} = \sum_{i=1}^{N} p_i \bar{Y}_{N_i}
$$

(8.1.1)
where

\[ p_i = \frac{N_i}{N} \quad \text{and} \quad \bar{y}_{N_i} = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}. \]

It is known that for simple random sampling the sample mean is an unbiased estimator of the population mean. Here

\[ \bar{y}_{st} = \sum_{i=1}^{k} p_i \bar{y}_{n_i} \]

where

\[ \bar{y}_{n_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad (8.1.2) \]

is an estimator of the population mean and its variance is given by

\[ \text{Var}(\bar{y}_{st}) = \sum_{i=1}^{k} p_i^2 \left( \frac{1}{n_i} - \frac{1}{N_i} \right) S_i^2 \]

where

\[ S_i^2 = \frac{1}{N_i-1} \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_{N_i})^2 \]

8.2. Optimum Allocation of Sample Size.

In the theory of sampling techniques the results developed are purely concerned with the particular finite population under consideration. Consider a finite population which is changing with time. Here the knowledge of stochastic process which has engendered it may be useful. The actual finite population
with which the Statistician is dealing with can be regarded as one particular set of realization of infinite values. Cochran (1946) was the first to visualize the process by assuming the particular finite population to be a random sample from a super-population. Rao (1968) has employed this super-population model of Cochran for allocating sample sizes in stratified sampling.

Recently, the model suggested for finite population by Cochran (1946) has been further developed by Avadhani and Sukhatme (1970). Commenting on their model Avadhani and Sukhatme (1970) have rightly remarked that their model is confined solely to the finite population under consideration and no assumptions of any distributional form are involved. Tacitly, in sampling from finite populations, this approach appears to be more comprehensible.

Consider the finite population model furnished by Avadhani and Sukhatme (1970).

\[
\begin{align*}
Y_{ij} &= \beta X_{ij} + e_{ij}, \quad i=1,2,\ldots,k, \quad j=1,2,\ldots,N_i \\
\sum_{j=1}^{N_i} e_{ij} &= 0 = \sum_{j=1}^{N_i} e_{ij}X_{ij} \\
\frac{e_{ij}^2}{X_{ij}^g} &= r, \quad r > 0 \quad 0 \leq g \leq 2
\end{align*}
\]
Under the finite population model, $\text{Var}(\bar{y}_{st})$ reduces to

$$
\text{Var}(\bar{y}_{st}) = \sum_{i=1}^{k} \left( \frac{1}{n_i} - \frac{1}{N_i} \right) p_i^2 \left( \beta S_{ix}^2 + \frac{r - x}{N_i - 1} \sum_{j=1}^{N_i} x_{ij} \right)
$$

where

$$
S_{ix}^2 = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_{N_i})^2,
$$

and

$$
\bar{x}_{N_i} = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}.
$$

The cost function may be noted as

$$
C = \sum_{i=1}^{k} c_i n_i
$$

where $c_i$ is the cost per unit in the $i$th stratum. When $c_i$ is the same for all strata, say $c$, then the cost function is

$$
C = cn
$$

Considering the cost function represented by equation (8.2.3) we determine the optimum values of $n_i$, which are given by

$$
n_i = \frac{p_i (\beta S_{ix}^2 + \frac{r - x}{N_i - 1} \sum_{j=1}^{N_i} x_{ij})^{1/2}}{\sqrt{\mu c_i}}
$$

where $\mu$ is some constant.
Now considering the fixed cost \( C_0 \), which is the budget amount within which it is desired to estimate the mean with the maximum precision. We get

\[
n_i = \frac{p_i (\beta^2 S^2_{ix} + \frac{r}{N_i - 1} \sum_{j=1}^{N_i} x_{ij})^{1/2} C_0}{\sqrt{c_i} \sum_{i=1}^{k} p_i \sqrt{c_i} (\beta^2 S^2_{ix} + \frac{r}{N_i - 1} \sum_{j=1}^{N_i} x_{ij}^2)^{1/2}}
\]

This is called the optimum allocation.

Now

\[
\beta^2 S^2_{ix} + \frac{r}{N_i - 1} \sum_{j=1}^{N_i} x_{ij}^2
\]

\[
= (\beta^2 + r) S^2_{ix} + \frac{rN_i}{N_i - 1} \left[ \frac{2}{\bar{x}^2_{N_i}} - \frac{1}{N_i} \left( \sum_{j=1}^{N_i} x_{ij}^2 - \sum_{j=1}^{N_i} x_{ij} \right) \right]
\]

Assuming \( \frac{N_i}{N_i - 1} \approx 1 \), and if \( S^2_{ix} \) are proportional to

\[
\frac{\bar{x}^2_{N_i} - \frac{1}{N_i} \left( \sum_{j=1}^{N_i} x_{ij}^2 - \sum_{j=1}^{N_i} x_{ij} \right)}{\bar{x}^2_{N_i} - \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}^2}, \] i.e. the square of the corrected coefficients of variation (c.c.v.) of the \( x \)-characteristic defined by \( \frac{S^2_{ix}}{\bar{x}^2_{N_i} - \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}^2} \), where \( \delta_i = N_i (\sum_{j=1}^{N_i} x_{ij}^2 - \sum_{j=1}^{N_i} x_{ij}^2) \)

are equal in all the strata, which is the case in many sample surveys encountered in practice. Under this condition the optimum allocation is given by
\[ n_i = \frac{-(\tau_i^2 - \delta_i)^{1/2}}{\sqrt{\sigma_i} \sum_{i=1}^{k} \sqrt{\sigma_i} (\tau_i^2 - \delta_i)^{1/2}} \]  \hspace{1cm} (8.2.7) \\

where \\

\[ T_i = \sum_{j=1}^{N_i} x_{ij}, \]

and the total sample size required for estimating the population mean with maximum precision for a fixed cost \( C_0 \) is given by

\[ n = \frac{C_0 \sum_{i=1}^{k} \left[ (T_i^2 - \delta_i)^{1/2} / \sqrt{\sigma_i} \right]}{\sum_{i=1}^{k} \sqrt{\sigma_i} (T_i^2 - \delta_i)^{1/2}} \]  \hspace{1cm} (8.2.8) \\

Thus we have the following theorem.

**Theorem 8.2.1.** Under the finite population model optimum allocation reduces to the allocation proportional to \( \frac{T_i^2 - \delta_i}{N_i} \) \( \frac{1/2}{\sqrt{\sigma_i}} \), where \( T_i = \sum_{j=1}^{N_i} x_{ij} \) is the total of the auxiliary variate \( X \) for the \( i \)th stratum, \( c_i \) the cost per unit in the \( i \)th stratum and \( \delta_i = N_i \left( \sum_{j=1}^{N_i} x_{ij}^2 - \sum_{j=1}^{N_i} x_{ij} \right) \), when the corrected coefficients of variations for the \( X \)-characteristic are equal in all the strata. For \( g = 2 \), the above theorem reduces to,

**Theorem 8.2.2.** For \( g = 2 \), under the finite population model, optimum allocation reduces to the allocation proportional to
\( \frac{T_i}{\sqrt{c_i}} \) where \( T_i = \sum_{j=1}^{N_i} x_{ij} \), is the total of the auxiliary variate \( X \) for the \( i \)th stratum and \( c_i \) the cost per unit in the \( i \)th stratum, when the coefficients of variation for the \( X \)-characteristic are equal in all the strata.

**Proof:** Under the finite population model, when \( g = 2 \), we have

\[
\delta_i = N_i \left( \sum_{j=1}^{N_i} x_{ij}^2 - \sum_{j=1}^{N_i} x_{ij}^g \right) = 0,
\]

Hence

\[
n_i = \frac{C_0 T_i}{\sqrt{c_i}} \sum_{i=1}^{k} \frac{T_i}{\sqrt{c_i}} \quad (8.2.9)
\]

Now when \( c_i = c \), i.e., the cost per unit in the \( i \)th stratum is the same for all the strata, then we have from equations (8.2.7) and (8.2.8)

\[
n_i = n \frac{(T_i^2 - \delta_i)^{1/2}}{\sum_{i=1}^{k} (T_i^2 - \delta_i)^{1/2}} \quad (8.2.10)
\]

Thus we have the following theorem,

**Theorem 8.2.3:** Under the finite population model, Neyman's optimum allocation reduces to allocation proportional to

\[
(T_i^2 - \delta_i)^{1/2}, \quad \text{where} \quad T_i = \sum_{j=1}^{N_i} x_{ij} \quad \text{is the total of the auxiliary} \]

variate $X$ for the $i$th stratum and $\delta_i = N_i \left( \sum_{j=1}^{N_i} x_{ij}^2 - \frac{\sum_{j=1}^{N_i} x_i}{N_i} \frac{g_i}{N_i} \right)$, when the corrected coefficients of variation for the $X$-characteristic are equal in all the strata.

For $g = 2$, the above theorem reduces to,

**Theorem 8.2.4.** For $g = 2$, under the finite population model, Neyman's optimum allocation reduces to allocation proportional to stratum totals of the auxiliary variate $X$, when coefficients of variation of $X$-characteristic are equal in all the strata.

**Proof:** Under finite population model, when $g = 2$, we have $\delta_i = 0$. Hence

$$n_i = n \frac{T_i}{\sum_{i=1}^{k} T_i}$$

Now when the population mean is to be estimated with a given variance $V_0$ at minimum cost. We have

$$\sum_{i=1}^{k} \left( \frac{1}{n_i} - \frac{1}{N_i} \right) p_i^2 \left( \beta^2 S_i^2 + \frac{r}{N_i-1} \sum_{j=1}^{N_i} x_{ij} \frac{g_i}{N_i} \right) = V_0.$$  

Substituting for $n_i$ from equation (8.2.5) and solving we get

$$n_i = \frac{\sum_{i=1}^{k} \left( \frac{1}{n_i} - \frac{1}{N_i} \right) p_i^2 \left( \frac{2x_{N_i}^2 - \delta_i}{N_i^2} \right) \sum_{i=1}^{k} \frac{1}{n_i} \left( \frac{2x_{N_i}^2 - \delta_i}{N_i^2} \right)^{1/2}}{\sqrt{c_i} \left[ V_0 + \frac{1}{N} \sum_{i=1}^{k} p_i \left( \frac{2x_{N_i}^2 - \delta_i}{N_i^2} \right) \right]}$$

$i = 1, 2, \ldots, k.$
Thus the minimum sample size required for estimating the mean with fixed variance $V_0$, under optimum allocation is given by

$$n = \frac{\sum_{i=1}^{k} p_i \left( \frac{\bar{x}_{N_1}^2}{N_1} - \frac{\delta_i}{N_1^2} \right)^{1/2} \sqrt{\sigma_i}}{V_0 + \frac{1}{N} \sum_{i=1}^{k} p_i \left( \frac{\bar{x}_{N_1}^2}{N_1} - \frac{\delta_i}{N_1^2} \right)^{1/2}}$$  \hspace{1cm} (8.2.11)$$

Putting $c_i = c$, we find the minimum sample size required for estimating the mean with fixed variance $V_0$ under Neyman's optimum allocation is given by

$$n = \frac{\left[ \sum_{i=1}^{k} p_i (\frac{\bar{x}_{N_1}^2}{N_1} - \frac{\delta_i}{N_1^2})^{1/2} \right]^2}{V_0 + \frac{1}{N} \sum_{i=1}^{k} p_i (\frac{\bar{x}_{N_1}^2}{N_1} - \frac{\delta_i}{N_1^2})}$$ \hspace{1cm} (8.2.12)$$

when $g = 2$, the above equation (8.2.12) reduces to,

$$n = \frac{k \left( \sum_{i=1}^{k} p_i \bar{x}_{N_1}^2 \right)^2}{V_0 + \frac{1}{N} \sum_{i=1}^{k} p_i \bar{x}_{N_1}^2}$$ \hspace{1cm} (8.2.13)$$

8.3. Illustration.

We illustrate the foregoing theory by considering the data furnished by Cochran (1963) which gives the number of inhabitants (in thousands), of 64 large cities in the United States, in 1920, grouped into two strata.
Table 8.3.1
Sizes of 64 cities (in thousands) in 1920, size \( (x_{ij}) \)

<table>
<thead>
<tr>
<th>Stratum 1</th>
<th>Stratum 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>797</td>
<td>314</td>
</tr>
<tr>
<td>773</td>
<td>298</td>
</tr>
<tr>
<td>748</td>
<td>296</td>
</tr>
<tr>
<td>734</td>
<td>258</td>
</tr>
<tr>
<td>588</td>
<td>256</td>
</tr>
<tr>
<td>577</td>
<td>243</td>
</tr>
<tr>
<td>507</td>
<td>238</td>
</tr>
<tr>
<td>507</td>
<td>237</td>
</tr>
<tr>
<td>457</td>
<td>235</td>
</tr>
<tr>
<td>438</td>
<td>235</td>
</tr>
<tr>
<td>415</td>
<td>216</td>
</tr>
<tr>
<td>401</td>
<td>208</td>
</tr>
<tr>
<td>387</td>
<td>201</td>
</tr>
<tr>
<td>381</td>
<td>192</td>
</tr>
<tr>
<td>384</td>
<td>180</td>
</tr>
<tr>
<td>315</td>
<td>179</td>
</tr>
</tbody>
</table>

Table 8.3.2

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Stratum size ( N_1 )</th>
<th>Stratum totals ( \sum_{j=1}^{N_1} x_{ij} )</th>
<th>( S_i )</th>
<th>Allocation proportional to ( S_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( T_i )</td>
<td></td>
<td></td>
<td>( T_i )</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>8349</td>
<td>4756619</td>
<td>163.3</td>
</tr>
<tr>
<td>2</td>
<td>48</td>
<td>7941</td>
<td>1474871</td>
<td>58.5</td>
</tr>
<tr>
<td>Stratum</td>
<td>g</td>
<td>$\delta_i$</td>
<td>$\sqrt{T_i^2 - \delta_i}$</td>
<td>c.c.v.</td>
</tr>
<tr>
<td>---------</td>
<td>---</td>
<td>------------</td>
<td>----------------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>1.7</td>
<td>64862736.0</td>
<td>2200.69</td>
<td>.1187</td>
</tr>
<tr>
<td>2</td>
<td>1.7</td>
<td>56201145.6</td>
<td>2618.84</td>
<td>.1073</td>
</tr>
<tr>
<td>1</td>
<td>1.8</td>
<td>54855360.0</td>
<td>3853.62</td>
<td>.6780</td>
</tr>
<tr>
<td>2</td>
<td>1.8</td>
<td>46120881.6</td>
<td>4115.65</td>
<td>.6829</td>
</tr>
<tr>
<td>1</td>
<td>1.9</td>
<td>36976704.0</td>
<td>5720.93</td>
<td>.4567</td>
</tr>
<tr>
<td>2</td>
<td>1.9</td>
<td>29024880.0</td>
<td>5833.91</td>
<td>.4817</td>
</tr>
</tbody>
</table>