WILTON DE OLIVEIRA BUSSAB

Departamento de Estatística - IME

U.S.P.

SÔBRE A ESTIMAÇÃO DA MÉDIA OU TOTAL EM SUBPOPULAÇÕES

TRABALHO APRESENTADO AO INSTITUTO DE MATEMÁTICA E ESTATÍSTICA DA UNIVERSIDADE DE SÃO PAULO, PARA MESTRADO EM ESTATÍSTICA APLICADA.

ABRIL 1971

PREFÁCIO

A teoria da amostragem baseia a grande maioria das suas conclusões na fixação a priori do tamanho da amostra a ser colhida. Mas, em alguns esquemas êsse tamanho não pode ser prefixado, sendo uma variável aleatória.

A maioria dos livros não trazem um tratamento sistemático do assunto. O nosso propósito foi fazer um levantamento da literatura sôbre êsse assunto, procurando sistematizar o tratamento, bem como desenvolver sugestões dadas pelos livros e não desenvolvidas. Em outros casos procuramos verificar a possibilidade de aplicação de certos estimadores desenvolvidos em casos gerais, ao nosso específico problema.

Dêsse modo dividimos o trabalho em três partes:

CAPÍTULO I - Introdução ao problema, às notações, e abordando o caso de amostragem casual simples. Inicialmente vemos algumas maneiras na estimação da média, procurando comparar os efeitos de cada estimador, e em seguida tratamos da estimativa do total;

CAPÍTULO II - Fazemos a mesma abordagem para a amostragem estratificada;

CAPÍTULO III - Uma rápida apresentação para a amostragem em dois estágios.

É nosso interêsse pesquisar no futuro, outros estimadores subpopulacionais bem como em outros esquemas de amostragem.

Finalmente queremos agradecer à tôdos aqueles que diretamente ou indiretamente nos ajudaram na elaboração dêste trabalho. Principalmente ao Professor Doutor Lindo Fava, nosso orientador, sugerindo o tema dêste trabalho bem como nos elucidando várias dúvidas; e ao senhor João Baptista Esteves de Oliveira, pelo trabalho de datilografia.

São Paulo, abril de 1971

CAPÍTULO I -

1-INTRODUÇÃO

Consideremos uma população N, com N unidades de amostragem, e em relação a qual está associada uma variável Y. Consideremos ainda imersa em N uma subpopulação M, isto é, algumas unidades da população possuem em comum determinada característica que as faz pertencer a essa subpopulação. O que acontece frequentemente é estarmos interessados na estimação de parâmetros desta subpopulação. Vejamos alguns exemplos:

(i) Quer-se estimar a renda média em domicílios que tenham mais de três adultos. Geralmente não dispomos da listagem dos domicílios com esta particular especificação, "mais de três adultos", mas a listagem de tôdos os domicílios. A identificação *a priori* das unidades amostrais da subpopulação pode ser muito cara ou difícil,o que nos impede de trabalhar apenas com a enumeração das unidades para a subpopulação, e temos que usar a listagem toda no processo de seleção da amostra.

(ii) Numa amostragem realizada estimou-se determinado parâmetro para uma população de pessoas adultas. Usando a mesma amostra quer-se estimar o mesmo parâmetro, mas agora restrita apenas à subpopulação masculina.

(iii) Uma grande emprêsa tem arquivado os canhotos de cheques emitidos no último ano, deseja-se estimar através de uma amostra o total em cruzeiros pagos com cheques de valor inferior a determinada quantia. Dificilmente dispõe-se da enumeração dos cheques nestas condições, e precisamos usar novamente a listagem de tôdos os cheques para côlhêr a amostra.

Dos exemplos acima, constatamos que a grande dificuldade na estimação de parâmetros da subpopulação,é o desconhecimento da enumeração das unidades amostrais da mesma. Êste fato sugere-nos dois caminhos a seguir para determinação da amostra;

-1º) Vamos selecionando elementos da população até encontrarmos um número fixado de elementos da subpopulação, abandonando os elementos não pertencentes a ela, e que foram selecionados.

-2º) Sorteamos uma amostra de tamanho fixado da população N, e desta amostra consideramos apenas os elementos pertencentes à subpopulação.

O primeiro procedimento leva-nos a uma amostragem casual simples de tamanho fixo. Mas podemos estar sujeito a colhêr um número

muito grande de elementos da população até atingirmos o número pré-fixado de elementos da subpopulação.

O segundo procedimento embora fixando o número de elementos de N que comporão a amostra, traz o inconveniente de que o número de elementos de M encontrados naquela amostra é uma variável aleatória.

Neste trabalho iremos tratar,principalmente, do uso do segundo processo, apresentado acima, para a estimação de parâmetros da subpopulação.

Embora usemos a palavra subpopulação, convém lembrar que o mesmo conceito aparece na literatura especializada sob outros títulos como: domínio, setor, subclasse, etc.

2.DEFINIÇÕES E NOTAÇÕES

Indicaremos a população N, do seguinte modo:

$$N = \{U_1, U_2, \ldots, U_N\}$$

Os possíveis valôres da variável Y, associada a esta população, serão indicados por:

$$Y_1, Y_2, \ldots, Y_N$$

Para identificarmos os valôres de Y correspondentes aos elementos da subpopulação, definiremos a variável auxiliar W do seguinte modo:

$$W_i = \begin{cases} Y_i & \text{se} \quad U_i \in M \\ 0 & \text{se} \quad U_i \notin M \end{cases} \qquad (2.1)$$

Então, os principais parâmetros da subpopulação, serão:

- valor médio da subpopulação

$$\bar{W} = \frac{1}{M} \sum_{i=1}^{N} W_i \qquad (2.2)$$

- total da subpopulação

$$W = \sum_{i=1}^{N} W_i = M\bar{W} \qquad (2.3)$$

onde M indica o número de elementos da subpopulação.

Colhida uma amostra de n elementos da população queremos usar estes elementos para estimar(2.2)e(2.3). A nossa amostra

$$y_1, y_2, \ldots, y_n$$

terá um número m de elementos da subpopulação, que iremos identificar do seguinte modo:

$$w_i = \begin{cases} y_i & \text{se} \quad y_i \in M \\ 0 & \text{se} \quad y_i \notin M \end{cases} \qquad (2.4)$$

Nestas condições, se quisermos estimar a média $\bar{W}$ de M, através de:

$$\bar{w} = \frac{1}{m} \sum_{i=1}^{n} w_i \qquad (2.5)$$

não podemos usar os teoremas básicos de amostragem (Cochran, teoremas 2.1, 2.2 e 2.4), porque m não é um número fixo mas varia de amostra para amostra, ou seja, m é uma variável aleatória discreta assumindo valôres de zero até n.

Neste trabalho, iremos ver algumas técnicas usadas na estimação de $\bar{W}$, e do total W da subpopulação. Além disso, quando nada se disser ao contrário, estamos considerando que a amostra foi tomada mediante amostragem casual simples, isto é, a probabilidade de qualquer unidade amostral ser incluida na amostra é constante e igual a $1/N$.

3. ESTIMAÇÃO COM m FIXO

Kish (pag.38) sugere como a primeira abordagem do problema, que em vez de especificarmos o tamanho n da amostra para a população, fixemos o tamanho m para a subpopulação. Em seguida vamos sorteando elementos de N, até completar os m diferentes elementos de M, e desprezando aqueles que não pertencem a esta subpopulação. Êsse procedimento, leva-nos a uma amostragem casual simples de m elementos, retirados de uma população de tamanho M.

De acôrdo com os teoremas usuais dêste tipo de amostragem, tiramos as seguintes conclusões:

- o estimador definido em (2.5)

$$\bar{w} = \frac{1}{m} \sum_{i=1}^{n} w_i = \frac{1}{m} \sum^{m} w_i \quad (*) \qquad (3.1)$$

é um estimador não viesado, e cuja variância é

$$V(\bar{w}) = (1-f) \frac{S_w^2}{m} \qquad (3.2)$$

onde

$$S_w^2 = \frac{1}{M-1} \sum^{M} (w_i - \bar{w})^2 \qquad (3.3)$$

(*) Usaremos o símbolo $\sum^{m} w_i$, para indicar a soma estendida apenas aos valôres não nulos da variável w_i na amostra.
$\sum^{M} W_i$, terá o mesmo significado para a população.

e

$$f' = \frac{m}{M} \tag{3.4}$$

e onde podemos usar como estimador de S_w^2, a variância da amostra:

$$s_w^2 = \frac{1}{m-1} \sum^{m} (w_i - \bar{w})^2 \tag{3.5}$$

Tal procedimento levou-nos à seguinte preocupação: qual o número médio de elementos sorteados até completar a amostra de tamanho desejado m?

Vamos dar uma resposta aproximada a êste problema, definindo inicialmente, a seguinte variável:

S_i: - número de elementos selecionados de N, após a escolha do i-ésimo elemento de M até a escolha do (i + 1)-ésimo elemento de M. (Para facilitar a linguagem, diremos que o processo está na i-ésima etapa).

Então, o número de elementos selecionados até o final da i-ésima etapa, é a variável aleatória

$$T_i = S_0 + S_1 + \ldots + S_{i-1} \tag{3.6}$$

Determinamos a distribuição de S_i. Ao iniciar-se a i-ésima etapa, já foram removidos T_{i-1} elementos de M, e teremos na população e subpopulação, respectivamente, os seguintes números de elementos:

$$\begin{aligned} A &= N - T_{i-1} \\ B &= M - i \end{aligned} \tag{3.7}$$

A probabilidade(*) de que na etapa i, sejam removidos k elementos, é dada pela probabilidade de escolhermos os primeiros (k - 1) elementos não pertencentes a M, vêzes a probabilidade do k-ésimo elemento pertencer a M. Então:

$$P(S_i=k) = \frac{(A-B)}{A} \cdot \frac{(A-B-1)}{A-1} \cdots \frac{(A-B-k+2)}{A-k+2} \cdot \frac{B}{A-k+1} = \frac{(A-B)!}{(A-B-k+1)!} \cdot \frac{(A-k)!}{A!} \cdot B =$$

$$= \frac{(A-B)!}{(A-B-k+1)!} \cdot \frac{(A-k)!}{A!} \cdot \frac{B!}{(B-1)!} = \frac{(A-k)!}{(A-k-B+1)!(B-1)!} \cdot \frac{(A-B)!B!}{A!}$$

$$P(S_i=k) = \frac{\binom{A-k}{B-1}}{\binom{A}{B}} \tag{3.8}$$

(*) Essa probabilidade é calculada, condicionada a um particular valor da variável T_{i-1}, isto é, supondo T_{i-1} fixo.

onde, k = 1, 2, 3, ..., A - B + 1. Como

$$1 = \sum_{k=1}^{A-B+1} P(S_i=k) = \sum \frac{\binom{A-k}{B-1}}{\binom{A}{B}} = \frac{1}{\binom{A}{B}} \sum \binom{A-k}{B-1}$$

vem imediatamente que:

$$\sum_{k=1}^{A-B+1} \binom{A-k}{B-1} = \binom{A}{B} \qquad (3.9)$$

Mas

$$E(S_i) = \sum_{k=1}^{A-B+1} k \cdot P(S_i=k) = \frac{1}{\binom{A}{B}} \sum k \binom{A-k}{B-1} \qquad (3.10)$$

usando a igualdade k = (A-B+1)-(A-B-k+1), em (3.10), vem

$$E(S_i) = (A-B+1) - \frac{1}{\binom{A}{B}} \sum_{k=1}^{A-B} (A-B-k+1) \frac{(A-k)!}{(A-B-k+1)!(B-1)!} =$$

$$= (A-B+1) - \frac{B}{\binom{A}{B}} \sum_{k=1}^{A-B} \binom{A-k}{B} = (A-B+1) - \frac{B}{\binom{A}{B}} \binom{A}{B+1}$$

esta última igualdade decorre da expressão (3.9). Simplificando os binomiais e reduzindo ao mesmo denominador chegamos a

$$E(S_i) = \frac{A + 1}{B + 1}$$

ou substituindo por (3.7)

$$E(S_i) = \frac{N - T_{i-1} + 1}{M - i + 1} \qquad (3.11)$$

O maior valor para $E(S_i)$, é aquêle obtido quando nas etapas anteriores 0, 1, 2, ..., (i-1), o primeiro elemento sorteado já pertence a M, isto é, $S_0 = 1$, $S_1 = 1$, ..., $S_{i-1} = 1$. Acontecendo isto na etapa i, teremos a maior quantidade possível de elementos não pertencentes a M, e consequentemente $T_i = i$. Então,

$$E(S_i) = \frac{N - T_{i-1} + 1}{M - i + 1} \leq \frac{N - i + 1}{M - i + 1} \qquad (3.12)$$

Queremos determinar o número de elementos removidos até completar a amostra de tamanho m desejado, que é dado por

$$T_m = S_0 + S_1 + \ldots + S_{m-1} \qquad (3.13)$$

logo:

$$E(T_m) = E(S_0) + E(S_1) + \ldots + E(S_{m-1}) \qquad (3.14)$$

das considerações anteriores obtemos:

$$E(T_m) \leq \frac{N+1}{M+1} + \frac{N}{M} + \frac{N-1}{M-1} + \ldots + \frac{N-m+2}{M-m+2} =$$

$$= \frac{M+D+1}{M+1} + \frac{M+D}{M} + \ldots + \frac{M+D-m+2}{M-m+2} =$$

$$= m + D\,\frac{1}{M+1} + \frac{1}{M} + \frac{1}{M-1} + \ldots + \frac{1}{M-m+2} <$$

$$< m + D\,\frac{1}{M-m+2} + \frac{1}{M-m+2} + \ldots + \frac{1}{M-m+2} =$$

$$= m + \frac{m.D}{M-m+2} = m\,1 + \frac{D}{M-m+2}$$

onde $D = N - M$. Substituindo e simplificando

$$E(T_m) < m\ \frac{N-m+2}{M-m+2}$$

por outro lado o limite inferior de $E(T_m)$ é o valor $\underline{m}$, logo

$$m < E(T_m) < m\ \frac{N-m+2}{M-m+2} = m\ \ 1 + \frac{Q}{(1-f')P} \qquad (3.15)$$

onde

$$P = \frac{M}{N} \quad e \quad Q = 1 - P \qquad (3.16)$$

A expressão (3.15) nos informa que quanto menor fôr P, maior será o extremo superior do número médio esperado de unidades removidas.

Como o custo de uma amostragem é fundamental na escolha do esquema a ser usado, vamos determinar uma expressão que nos informe sôbre o custo total do esquema sugerido nesta secção.

Suponhamos que para investigar uma unidade amostrada gasta-se $\underline{a}$ cruzeiros. Caso a unidade pertença a subpopulação, gasta-se mais $\underline{b}$ cruzeiros para completar a investigação. Assim a variável C, custo total, é dada por

$$C = C_0 + mb + aT_m$$

onde C_o é o custo fixo. Então,

$$E(C) = C_0 + mb + a\,E(T_m)$$

Esta última expressão combinada com (3.15), nos dá subsídios para analisar o custo esperado da pesquisa.

4. AMOSTRA "CONDICIONAL" DE m ELEMENTOS

Cochran (pag.33), quando aborda inicialmente o problema de subpopulações, e Kish (pag.42), sugerem um método para estimar $\bar{W}$. Consideram a distribuição de $\bar{w}$ sôbre as amostras de n elementos da população, nas quais aparecem exatamente m elementos da subpopulação. Isto é, dentre tôdas as possíveis amostras que podemos formar

de tamanho n, determinamos a distribuição condicional de $\bar{w}$ restrita àquelas que possuam exatamente m elementos de M.

O número de amostras nas condições acima é

$$\binom{N - M}{n - m} \cdot \binom{M}{m} \tag{4.1}$$

pois temos n lugares dos quais m devem ser preenchidos pelos M elementos da subpopulação, e os n-m restantes lugares pelos N-M elementos de M' (complementar de M em relação ao universo N). Dentre estas amostras o número daquelas que apresentam uma específica amostra de m unidades da subpopulação é

$$\binom{N - M}{n - m} \tag{4.2}$$

pois colocada a específica configuração, sobram n-m posições a serem preenchidas por N-M elementos. Então, a probabilidade de que seja escolhida uma particular amostra de m elementos de M é dada por

$$\frac{\binom{N - M}{n - m}}{\binom{N - M}{n - m}\binom{M}{m}} = \frac{1}{\binom{M}{m}}$$

O que equilave a afirmar que estamos fazendo uma amostragem casual simples, de m elementos de uma população de tamanho M. Se queremos estimar $\bar{W}$, podemos usar o estimador $\bar{w}$ definido em (2.5), e que está nas mesmas condições da secção anterior, é um estimador não viesado, cuja variância é dada por (3.2)

$$V(\bar{w}) = \frac{1 - f'}{m} S_w^2 \tag{4.3}$$

Frequentemente, desconhecemos a grandeza M, para calcularmos f'. Uma aproximação é substituir f' por

$$f = \frac{n}{N} \tag{4.4}$$

sendo esta uma aproximação razoável já que

$$E(f') = E\frac{m}{M} = \frac{n}{N} = f \tag{4.5}$$

Para demonstrar esta última afirmação introduziremos a variável auxiliar X, do seguinte modo:

$$X_i = \begin{cases} 1 \text{ se o elemento pertence a } M \\ 0 \text{ nos demais casos} \end{cases} \tag{4.6}$$

Usaremos x_i para indicar a mesma variável mas agora em relação a amostra. Assim

$$\sum_{i=1}^{N} X_i = M \quad \text{e} \quad \sum_{i=1}^{n} x_i = m \tag{4.7}$$

então:

$$E\left(\frac{m}{M}\right) = \frac{1}{M} E\left(\sum_{i=1}^{n} x_i\right)_N = \frac{1}{M} \sum_{i=1}^{n} E(x_i) \qquad (4.8)$$

mas,

$$E(x_i) = \bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{M}{N} \qquad (4.9)$$

substituindo em (4.8), vem:

$$E\left(\frac{m}{M}\right) = \frac{1}{M} \sum_{i=1}^{n} \frac{M}{N} = \frac{1}{M} \cdot m \cdot \frac{M}{N} = \frac{n}{N} \qquad (4.10)$$

o que demonstra (4.5).

Substituindo-se (4.4) e (3.5) em (4.3), temos que a variância de $\bar{w}$ "condicionada" ao encontro de $\underline{m}$ elementos da subpopulação é dada por:

$$V(\bar{w}) = \frac{1-f}{m} \cdot s_w^2 = \left(1-\frac{n}{N}\right) \cdot \frac{1}{m} \cdot \frac{1}{m-1} \sum_{i=1}^{m} (w_i - \bar{w})^2 \qquad (4.11)$$

5. $\bar{w}$ COMO ESTIMADOR RAZÃO

Usando as variáveis W e X, definidas por (2.1) e (4.6), respectivamente verificamos as seguintes relações:

$$\bar{W} = \frac{1}{M} \sum_{i=1}^{N} W_i = \frac{\sum_{i=1}^{N} W_i}{\sum_{i=1}^{N} X_i} = \frac{\bar{W}_o}{\bar{X}} = R \qquad (5.1)$$

onde:

$$\bar{W}_o = \frac{1}{N} \sum_{i=1}^{N} W_i \qquad (5.2)$$

e

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i = \frac{M}{N} \qquad (5.3)$$

bem como

$$\bar{w} = \frac{1}{m} \sum_{i=1}^{n} w_i = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{w}_o}{\bar{x}} = r \qquad (5.4)$$

onde

$$\bar{w}_o = \frac{1}{n} \sum_{i=1}^{n} w_i \qquad (5.5)$$

e

$$\bar{x} = \sum x_i = \frac{m}{n} \qquad (5.6)$$

Para determinarmos as propriedades dêste estimador, vamos usar a técnica usada por P.V.Sukahtme (sec. 4a.3 pag.139), para estimador razão geral, e particularizemos para a nossa situação.

Façamos:

$$w_i = \bar{W}_o + E_i \quad \text{o que resulta} \quad \bar{w}_o = \bar{W}_o + \bar{E} \tag{5.7}$$

onde

$$E(\bar{E}) = 0 \quad e \quad V(\bar{E}) = E(\bar{E}^2) = (1 - f).\frac{S_o^2}{n} \tag{5.8}$$

onde

$$S_o^2 = \frac{1}{N-1} \sum_{i=1}^{N} (W_i - \bar{W}_o)^2 \tag{5.9}$$

Do mesmo modo:

$$x_i = \bar{X} + E_i' \qquad \bar{x} = \bar{X} + \bar{E}' \tag{5.10}$$

com

$$E(\bar{E}') = 0 \quad e \quad V(\bar{E}') = (1-f)\frac{S_x^2}{n} \tag{5.11}$$

onde

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2 = \frac{M}{N-1}\left(1 - \frac{M}{N}\right) \tag{5.12}$$

Substituindo (5.7) e (5.10) em (5.4), teremos:

$$n = \frac{\bar{w}_o}{\bar{x}} = \frac{\bar{W}_o + \bar{E}}{\bar{X} + \bar{E}'} = \frac{\bar{W}_o}{\bar{X}} \cdot \frac{1 + \frac{\bar{E}}{\bar{W}_o}}{1 + \frac{\bar{E}'}{\bar{X}}} \tag{5.13}$$

Como n, é um quociente de duas variáveis aleatórias, vamos calcular a esperança aproximada, desenvolvendo:

$$\left(1 + \frac{\bar{E}'}{\bar{X}}\right)^{-1}$$

em série, mas para que a série seja convergente, precisamos ter

$$\left|\frac{\bar{E}'}{\bar{X}}\right| < 1$$

isto é

$$\left|\frac{\bar{x} - \bar{X}}{\bar{X}}\right| < 1 \rightarrow \left|\frac{\bar{x}}{\bar{X}} - 1\right| < 1 \rightarrow 0 < \frac{\bar{x}}{\bar{X}} < 2$$

ou substituindo

$$0 < \frac{\frac{m}{n}}{\frac{M}{N}} < 2$$

que resulta em:

$$\frac{m}{n} > 0 \quad e \quad \frac{m}{n} < 2\,\frac{M}{N}$$

A primeira desigualdade, quase sempre está satisfeita, porque o número de amostras de tamanho n que não contenha nenhum elemento da subpopulação, pode ser consideraconsiderado como uma raridade, e quanto maior fôr n, mais difícil fica não aparecer elementos de M. A segunda desigualdade já é mais delicada. Necessita que a fração da subpopulação em relação a população tôda na amostra, seja infe-

rior ao dôbro da mesma fração no todo. Como frequentemente as subpopulações são frações muito pequenas, esta desigualdade estará satisfeita num número reduzido de vêzes. Mas no caso em que subpopulação é grande mais que a metade da população, essa desigualdade se verifica sempre.

$$\left(1 + \frac{\bar{E}'}{\bar{X}}\right)^{-1} = 1 - \frac{\bar{E}'}{\bar{X}} + \left(\frac{\bar{E}'}{\bar{X}}\right)^2 - , \ldots \qquad (5.14)$$

Substituindo êste valôr em (5.13), teremos:

$$n = \frac{\bar{W}_o}{\bar{X}}\left(1 + \frac{\bar{E}}{\bar{W}_o}\right)\left\{1 - \frac{\bar{E}}{\bar{X}} + \left(\frac{\bar{E}'}{\bar{X}}\right)^2 + \ldots\right\}$$

$$= \frac{\bar{W}_o}{\bar{X}}\left\{1 + \frac{\bar{E}}{\bar{W}_o} - \frac{\bar{E}'}{\bar{X}} - \frac{\bar{E}'\bar{E}}{\bar{X}\bar{W}_o} + \left(\frac{\bar{E}'}{\bar{X}}\right)^2 + \left(\frac{\bar{E}'}{\bar{X}}\right)^2 \cdot \frac{-\bar{E}-}{\bar{W}_o} + \ldots\right\} \qquad (5.15)$$

Consideremos negligenciáveis a contribuição dos têrmos envolvendo $\bar{E}'$ e $\bar{E}$, com potências superior ao segundo grau, e calculando a esperança, teremos:

$$E(r) = E\left(R\left\{1 + \frac{\bar{E}}{\bar{W}_o} - \frac{\bar{E}'}{\bar{X}'} - \frac{\bar{E}'\bar{E}}{\bar{X}\bar{W}_o} + \left(\frac{\bar{E}'}{\bar{X}}\right)^2\right\}\right) =$$

$$= R\left\{1 + \frac{E(\bar{E})}{\bar{W}_o} - \frac{E(\bar{E}')}{\bar{X}'} - \frac{E(\bar{E}'\bar{E})}{\bar{X}\,\bar{W}_o} + \frac{E(\bar{E}'^2)}{\bar{X}^2}\right\}$$

mas $\quad E(\bar{E}) = 0 \,, \qquad E(\bar{E}') = 0$

$$E(\bar{E}'^2) = (1 - f)\frac{S_x^2}{n} \quad \text{segundo (5.11)}$$

$$E(\bar{E}'\bar{E}) = \frac{N - n}{N}\,\frac{1}{n}\,\rho\, S_n . S_o$$

esta última expressão encontra-se desenvolvida em Sukahtme.
Substituindo êstes valôres em E(r), encontramos:

$$E(r) = R.\left\{1 + \frac{(1-f)}{\bar{X}^2}\,\frac{S_x^2}{n} - \frac{(1-f)}{\bar{X}\bar{W}_o} \cdot \frac{\rho S_x S_o}{n}\right\} =$$

$$= R.\left\{1 + \frac{(1-f)}{n}\,\frac{S_x^2}{\bar{X}^2} - \frac{\rho S_x S_o}{\bar{X}\bar{W}_o}\right\} \qquad (5.16)$$

Para que seja um estimador não viesado, é preciso que a expressão entre parêntesis seja igual a 0 (zero), isto é,:

$$\frac{S_x^2}{\bar{X}^2} - \frac{\rho S_x S_o}{\bar{X}\bar{W}_o} = 0$$

$$\frac{\rho S_o}{\bar{W}_o} = \frac{S_x}{\bar{X}}$$

$$\bar{W}_o = \rho \frac{S_o}{S_x} \bar{X} \qquad (5.17)$$

Esta última expressão significa que a reta de regressão de W em função de X, passa pela origem. Para o nosso problema êste fato sempre se verifica, porque a todo elemento não pertencente a subpopulação M, está associado a observação (0,0). O que nos mostra que o estimador assim definido, é não viesado quando se usa a aproximação até segunda ordem. Então,

$$E(\bar{w}) = E(r) = R = \bar{W} \qquad (5.18)$$

Algumas vêzes é bastante importante o conhecimento da magnitude do verdadeiro viés, e isso pode ser calculado usando a sugestão dada por Hartley e Ross, como encontramos em Fava (pág.9). Indicando por ε o verdadeiro viés de $\bar{w}$, calculado como estimador razão, teremos:

$$\varepsilon = E(\bar{w}) - \bar{W} = -\{\bar{W} - E(\bar{w})\} = -\left[\frac{\bar{W}_o}{P} - E\left(\frac{\bar{w}_o}{\bar{x}}\right)\right] =$$

$$= -\left[\frac{\bar{W}_o}{\bar{X}} - E\left(\frac{\bar{w}_o}{\bar{x}}\right)\right] = -\frac{1}{\bar{X}}\left[\bar{W}_o - \bar{X}\, E\left(\frac{\bar{w}_o}{\bar{x}}\right)\right] =$$

$$= -\frac{1}{\bar{X}} E\left\{\left[\frac{\bar{w}_o}{\bar{x}} - E\left(\frac{\bar{w}_o}{\bar{x}}\right)\right]\left[\bar{x} - E(\bar{x})\right]\right\} = -\frac{1}{\bar{X}} \mathrm{Cov}\left(\frac{\bar{w}_o}{\bar{x}}, \bar{x}\right) \qquad (5.19)$$

onde $\mathrm{Cov}\left(\frac{\bar{w}_o}{\bar{x}}, \bar{x}\right)$ é a covariância entre as variáveis $\frac{\bar{w}_o}{\bar{x}}$ e $\bar{x}$. Então,

$$|\varepsilon| = \frac{1}{\bar{X}} \left|\mathrm{Cov}\left(\frac{\bar{w}_o}{\bar{x}}, \bar{x}\right)\right| \leq \frac{1}{\bar{X}} S\left(\frac{\bar{w}_o}{\bar{x}}\right) S_{\bar{x}} \qquad (5.20)$$

pois $|\rho| \leq 1$.

$$\left|\frac{\mathrm{Cov}\left(\frac{\bar{w}_o}{\bar{x}}, \bar{x}\right)}{S\left(\frac{\bar{w}_o}{\bar{x}}\right) S_{\bar{x}}}\right| \leq 1 \longrightarrow \left|\mathrm{Cov}\left(\frac{\bar{w}_o}{\bar{x}}, \bar{x}\right)\right| \leq S\left(\frac{\bar{w}_o}{\bar{x}}\right) S_{\bar{x}}$$

onde ρ é o coeficiente de correlação entre as duas varíaveis. Finalmente

$$|\varepsilon| \leq \frac{1}{\bar{X}} S_{\bar{w}} S_{\bar{x}} \quad \text{ou} \quad \frac{|\varepsilon|}{S_{\bar{w}}} \leq \frac{S_{\bar{x}}}{\bar{X}} \qquad (5.21)$$

mas como $S_{\bar{x}}$ é igual a S_x dividido por $\sqrt{n}$, e como já vimos em (5.12)

que $S_x = \sqrt{\frac{M}{N-1}\left(1 - \frac{M}{N}\right)} \cong \sqrt{P.Q}$

onde $P = \frac{M}{N}$ e $Q = 1 - P$, substituindo em (5.21) encontramos:

$$\frac{|\varepsilon|}{S_{\bar{x}}} \leq \frac{1}{n}\frac{P}{Q} \qquad (5.22)$$

e esta expressão nos fornece qual a intensidade do viés. Se a fração P fôr pequena, vemos que para manter o viés reduzido teremos necessidade de que a amostra seja bem grande. Vejamos um exemplo numérico. Suponhamos que queremos manter um viés relativo inferior à 0,1. Então devemos ter:

$$\sqrt{\frac{1}{n}\frac{Q}{P}} = 0,1 \longrightarrow \frac{1}{n}\frac{Q}{P} = 0,01$$

$$n = \frac{Q}{0,01P}$$

Suponhamos inicialmente que a fração da subpopulação é 80% da população, isto é, P = 0,8, então necessitamos n = 25 elementos, se P = 0,5 então o tamanho da amostra será 100 elementos, e se P é apenas 20% necessitaremos de 400 elementos na amostra. O que nos mostra que necessitamos de amostras muito grandes quando a proporção da subpopulação é pequena.

6. CALCULO DA VARIÂNCIA DE $\bar{w}$, COMO ESTIMADOR RAZÃO

Para calcular a esperança de $\bar{w}$, tivemos que fazer aproximações, no cálculo da variância também vamos fazê-las, mas em duas aproximações. A primeira é usando o desenvolvimento de r obtido em (5.15), apenas com potências até o primeiro grau. Esta é a técnica usada por Sukahtme (pág.146), para estimador razão no caso geral, e que particularizaremos para êste problema. Assim temos:

$$r = R\left\{1 - \frac{\bar{E}}{\bar{W}_o} + \frac{\bar{E}'}{\bar{X}}\right\} \qquad (6.1)$$

Mas do fato de $E(r) \cong R$, obtido na secção anterior, teremos:

$$V_1(\bar{w}) = V(r) = E\left[\{r - E(r)\}^2\right] \cong E\left[\{r-R\}^2\right] =$$

$$= R^2\left\{\frac{E(\bar{E}'^2)}{\bar{X}^2} + \frac{E(\bar{E}^2)}{\bar{W}_o^2} - \frac{2E(\bar{E}'\bar{E})}{\bar{X}\,\bar{W}_o}\right\}$$

Calculando as esperanças, segundo sugestão da secção anterior,

$$V_1(\bar{w}) = R^2\left\{\frac{(1-f)}{\bar{X}^2 n}S_x^2 + \frac{(1-f)}{\bar{W}_o^2\, n}S_o^2 - \frac{2(1-f)}{n}\frac{\rho S_x S_o}{\bar{X}\,\bar{W}_o}\right\} =$$

$$= R^2\left\{\frac{(1-f)}{n}\frac{S_x^2}{\bar{X}^2} + \frac{S_o^2}{\bar{W}_o^2} - \frac{2\rho S_x S_o}{\bar{X}\,\bar{W}_o}\right\}$$

mas de (5.17), tiramos imediatamente que:

$$V_1(\bar{w}) = R^2 \frac{(1-f)}{n} \left[\frac{S_o^2}{\bar{W}_o^2} + \frac{S_x^2}{\bar{X}^2} \right] \qquad (6.2)$$

Vamos tornar esta expressão mais simples, lembrando que:

$R = \bar{W}$ de (5.1)

$S_x^2 = \frac{M}{N-1}\left(1 - \frac{M}{N}\right)$ de (5.12)

$\bar{X} = \frac{M}{N}$ de (5.3)

e determinando as novas relações

$$\bar{W}_o = \frac{1}{N} \sum^{N} W_i = \frac{M}{N} \frac{1}{N} \sum^{M} W_i = \frac{M}{N} \bar{W} = P\bar{W} \qquad (6.3)$$

$$(N-1)S_o^2 \cong \sum^{N} (W_i - \bar{W}_o)^2 = \sum^{N} W_i^2 - N\bar{W}_o^2 = \sum^{N} W_i^2 - M\bar{W}^2 + M\bar{W}^2 - N\bar{W}_o^2 =$$

$$= \sum^{M} W_i^2 - M\bar{W}^2 + M\bar{W}^2 - N\frac{M^2}{N^2}\bar{W}^2 = (M-1)\, S_w^2 + M\bar{W}^2\, Q$$

isto é $S_o^2 \cong P\, S_w^2 + PQ\, \bar{W}^2$ (6.4)

Substituindo estes últimos resultados em (6.2), e simplificando temos:

$$V_1(\bar{w}) = \frac{(1-f)}{n} \frac{N}{M} S_w^2 = (1-f)\frac{S_w^2}{fM} \qquad (6.5)$$

Usa-se como estimador de S_w^2 a variância amostral

$$s_w^2 = \frac{1}{m-1} \sum^{m} (w_i - \bar{w})^2 \qquad (6.6)$$

Como segunda aproximação usaremos o processo desenvolvido por Hansen (pág. 114, 2º vol.),

$$V_2(\bar{w}) = V(r) \cong E\{[r - R]^2\}$$

e que através da propriedade de esperança condicional, pode ser escrita da seguinte maneira:

$$V_2(\bar{w}) = E\left[E_m (r - R)^2\right] \qquad (6.7)$$

onde E_m significa a esperança condicionada à ocorrência de $\underline{m}$ indivíduos da subpopulação.

$$E_m\left[(r - R)^2\right] = (1-f')\frac{S_w^2}{m} = \left(\frac{M-m}{M}\right)\frac{1}{m} S_w^2 = \left(\frac{1}{m} - \frac{1}{M}\right) S_w^2 \qquad (6.8)$$

Substituindo êsse resultado em (6.7), vem

$$V_2(\bar{w}) = E\left[\left(\frac{1}{m} - \frac{1}{M}\right) S_w^2\right] = S_w^2 \left[E\left(\frac{1}{m}\right) - \frac{1}{M}\right] \qquad (6.9)$$

De acôrdo com a definição de $\bar{E}'$ dada em (5.10), podemos escrever:

$$m = n\bar{x} = n(\bar{X} + \bar{E}') = n\bar{X}\left(1 + \frac{\bar{E}'}{\bar{X}}\right) \qquad (6.10)$$

mas $$\left(1 + \frac{\bar{E}'}{\bar{X}}\right)^{-1} = 1 - \frac{\bar{E}'}{\bar{X}} + \left(\frac{\bar{E}'}{\bar{X}}\right)^2 + \ldots$$

logo $$E\left(\frac{1}{m}\right) = \frac{1}{n\bar{X}} E\left(1 - \frac{\bar{E}'}{\bar{X}} + \left(\frac{\bar{E}'}{\bar{X}}\right)^2 + \ldots\right) \cong \frac{1}{n\bar{X}}\left(1 + \frac{E(\bar{E}'^2)}{\bar{X}^2}\right) =$$

$$= \frac{N}{nM}\left\{1 + \frac{(1-f)\cdot\frac{1}{N}\frac{M}{N-1}\left(1 - \frac{M}{N}\right)}{\frac{M^2}{N^2}}\right\} \qquad (6.11)$$

Substituindo em (6.9), vamos encontrar:

$$V_2(\bar{w}) = \left\{\frac{N}{n\,M}\left[1 + \frac{(1-f)}{n}\frac{N}{M}\left(1 - \frac{M}{N}\right)\frac{N}{N-1}\right] - \frac{1}{M}\right\} S_w^2 =$$

$$\cong \frac{N}{n\,M}\left\{(1-f)\left[1 + \frac{1}{n}\frac{Q}{P}\left(\frac{N}{N-1}\right)\right]\right\} S_w^2 =$$

$$\cong \frac{1-f}{f\,M}\left(1 + \frac{1}{n}\frac{Q}{P}\right) S_w^2 \qquad (6.12)$$

Na prática usamos s_w^2 definido em (6.6) como estimador de S_w^2; m como estimador de fM e $\frac{m}{n}$ como estimador de P.

7. COMPARAÇÃO DAS VARIÂNCIAS OBTIDAS PARA $\bar{w}$

Vamos procurar comparar as variâncias obtidas na secção anterior, com a variância de $\bar{w}$ obtida quando se supõe a escolha de uma amostra casual simples, de m elementos de uma população com M elementos, e como já foi visto em (3.2) é dada por

$$V(\bar{w}) = (1-f')\frac{S_w^2}{m} \qquad (7.1)$$

Se compararmos com a variância do estimador razão, obtidas em (6.5)

$$V_1(\bar{w}) = (1-f)\frac{S_w^2}{fM} \qquad (7.2)$$

vemos que as diferenças são mínimas, visto que, podemos supor $f \cong f'$, e $E(m) = fM$, e sabemos que m pode ser confundido com a sua esperança, o que nos leva a concluir que as diferenças nas duas formas são desprezíveis. Para aplicação, ambas as formas devem sofrer alteração quando se desconhece M. Para (7.1) devemos substiuir f' por f, e em (7.2) fM por m.

Comparamos agora com a variância obtida em (6.12), ou seja, como estimador razão em segunda aproximação,

$$V_2(\bar{w}) = (1-f)\left(1 + \frac{1}{n}\frac{Q}{P}\right)\frac{s_w^2}{fM} \qquad (7.3)$$

A variância neste caso é maior que a variância obtida pela fórmula (7.1), êsse acréscimo é devido à aleatoriedade de m na amostra. Analisemos um pouco mais detidamente o acréscimo

$$\frac{Q}{nP}\,\frac{s_w^2}{fM} \qquad (7.4)$$

cuja interpretação é bastante intuitiva:

i) Êsse acrescimo diminui quando o número de unidades amostradas aumenta;

ii) Quanto menor a fração da subpopulação em relação ao tamanho da população, maior é o acréscimo observado na variância.

8. ESTIMATIVA DO VALOR TOTAL W DAS SUBPOPULAÇÕES

Suponhamos agora que estamos interessados em estimar o total

$$W = \sum_{i=1}^{N} W_i = \sum^{M} W_i = M\bar{W}$$

Para isso vamos dividir o tratamento em três casos:

1. O total Y da população é conhecido;
2. O tamanho M da subpopulação é conhecido;
3. Y e M desconhecidos.

8.1-O total Y é conhecido

Existem algumas pesquisas em subpopulação, que por algum motivo, conhecemos o valor total da população. Como vimos no início deste trabalho, tinhamos uma população na qual observavamos uma variável Y, e através de um artifício criamos uma variável auxiliar W. Assim para cada unidade da população temos associado o par (W_i, Y_i), o que nos permite usar um estimador razão, na estimativa do total W. Analisando a relação entre a variável W e Y podemos nos decidir por qualquer um dos estimadores razão.

8.2-O tamanho M da subpopulação é conhecido

Neste caso a estimativa do total pode ser dado através de:

$$\tilde{w} = M\bar{w} \qquad (8.1)$$

Na qual podemos usar $\bar{w}$, como qualquer um dos estimadores propostos nas secções anteriores. Usando o estimador definido na secção 3, ou no caso da amostragem "condicional", visto na secção 4, $\tilde{w}$ é um estimador não viesado com a seguinte variância:

$$V(\tilde{w}) = \frac{1-f'}{m} M^2 S_w^2 \qquad (8.2)$$

de acôrdo com as conclusões das citadas secções.

Se usarmos $\bar{w}$ como o estimador razão da secção 6, e em concordância com as conclusões lá obtidas, temos em $\tilde{w}$ um estimador não viesado, até segunda ordem de aproximação, e cuja variância é dada, em primeira aproximação por (6.5)

$$V_1(\tilde{w}) = \frac{1-f}{f} M S_w^2 \qquad (8.3)$$

e em segunda aproximação, conforme (6.12)

$$V_2(\tilde{w}) = \frac{1-f}{f} M \left(1 + \frac{1}{n}\frac{Q}{P}\right) S_w^2 \qquad (8.4)$$

8.3-Y e M desconhecidos

Êste é o caso mais geral e mais frequente. Usaremos aquí o artifício encontrado em Cochran (pág.34).O estimador de W proposto é o obtido pela multiplicação do total observado na amostra pelo fator de crescimento $\frac{N}{n}$:

$$\hat{w} = \frac{N}{n} w \qquad (8.5)$$

onde

$$w = \sum_{i=1}^{n} w_i$$

então

$$\hat{w} = \frac{N}{n} \sum^{n} w_i = N\bar{w}_o \qquad (8.6)$$

$$E(\hat{w}) = N\, E(\tilde{\bar{w}}_o) = N\, \bar{W}_o = N \frac{\Sigma W_i}{N} = \sum W_i = W \qquad (8.7)$$

ou seja, êste é um estimador não viesado.

Da expressão (8.6) podemos calcular a variância rapidamente lembrando que $\bar{w}_o$, representa a média de uma amostra de n elementos, colhida de uma população com N elementos, e pelas conclusões da amostragem casual simples:

$$V(\bar{w}_o) = (1-f) \frac{S_o^2}{n} \qquad (8.8)$$

Substituindo em (8.6)

$$V(\hat{w}) = V(N\bar{w}_o) = N^2 V(\bar{w}_o) = N^2(1-f)\frac{S_o^2}{n} \qquad (8.9)$$

onde usamos

$$s_o^2 = \frac{1}{n-1}\{\sum(w_i - \bar{w}_o)^2\}$$

como estimador de S_o^2. O que pode causar um pouco de estranheza na última expressão, é o fato da variância ser calculada inclusive para os zeros da população, o que a torna bem maior do que aquela obtida dos valôres não nulos.

Algumas vêzes, é conveniente exprimir (8.9) em função de S_w^2. Usando o resultado (6.4) e (6.13), podemos escrever:

$$S_o^2 = \frac{M-1}{N-1} S_w^2 + \frac{M}{N-1}\left(1 - \frac{M}{N}\right)\bar{W}^2 \cong PS_w^2 + PQ\bar{W}^2 \qquad (8.10)$$

substituindo em (8.9)

$$V(\hat{w}) = \frac{N^2(1-f)}{n}\{PS_w^2 + PQ\bar{W}^2\} \qquad (8.11)$$

Como frequentemente desconhecemos S_w^2, P, Q e $\bar{W}$, um estimador de (8.11) é

$$v(\hat{w}) = (1-f)\frac{N^2}{n}\left[\frac{m}{n} s_w^2 + \frac{m}{n}\left(1 - \frac{m}{n}\right)\bar{w}^2\right]$$

8.4-Comparação das variâncias

Vamos comparar as variâncias obtidas quando se desconhece tôdas as informações da subpopulação com aquela obtida do conhecimento do tamanho da mesma. (Cochran - pág. 36).

Da expressão (8.2) temos que:

$$V(\tilde{w}) = (1-f')\frac{M^2}{m} S_w^2 \qquad (8.12)$$

onde m é uma variável aleatória, mas já vimos que E(m) = nP. Então, supondo m = nP, teremos:

$$f' = \frac{m}{M} = \frac{nP}{NP} = f$$

substituindo em (8.12)

$$V(\tilde{w}) = (1-f)\frac{N^2P^2}{nP} S_w^2 = (1-f)\frac{N^2P}{n} S_w^2 \qquad (8.13)$$

A variância quando se desconhece o tamanho da subpopulação é dada por (8.11):

$$V(\hat{w}) = (1-f)\frac{N^2}{n} P\{S_w^2 + Q\bar{W}^2\} \qquad (8.14)$$

então

$$\frac{V(\hat{w})}{V(\tilde{w})} = \frac{S_w^2 + Q\bar{W}^2}{S_w^2} = 1 + \frac{Q}{C_w^2} \qquad (8.15)$$

onde $C_w = \frac{S_w}{\bar{W}}$ é o coeficiente de variação da subpopulação. Evidentemente a variância é maior quando não se tem informações sôbre o tamanho da subpopulação. Em outras palavras, o lucro que se tem pelo conhecimento dêsse tamanho aumenta, quanto menor fôr a fração P,

8.5-Outra aplicação

As conclusões obtidas na secção anterior são de grande utilidade no seguinte problema (Kish, pág. 435). Suponhamos que queremos estimar o total Y, de uma população em que muito dos seus valôres são zeros. A pergunta que se coloca é a seguinte: De quanto podemos reduzir a variância se restringirmos nossa amostragem apenas aos valôres não nulos da população?

A variável Y da população N, pode ser separada em duas subpopulações a saber: M dos elementos em que Y não é nulo e M' dos zeros da população. Usaremos aqui novamente, a variável W definida em (2.1). Assim êste problema, é o mesmo dos tratados nas secções anteriores, ou seja, queremos comparar a variância entre dois tipos de estimadores,

$$\hat{w} = N\,\bar{w}_o = \frac{N}{n}\,w$$

e

$$\tilde{w} = M\,\bar{w} = \frac{M}{m}\,w$$

Aqui, ao contrário da secção (8.4), é conveniente exprimir as variâncias em função de S_o, que é a variância da população toda. Assim de (8.10), temos que

$$S_w^2 = \frac{S_o^2}{P} - Q\bar{W}^2 \qquad (8.16)$$

Substituindo em (8.2) teremos:

$$V(\tilde{w}) = \frac{(1-f')}{m} M^2 \left(\frac{S_o^2}{P} - Q\bar{W}^2 \right)$$

ou usando transformações anteriores

$$V(\tilde{w}) = \frac{(1-f)}{n\,P} N^2 P^2 \left\{ \frac{S_o^2 - PQ\bar{W}^2}{P} \right\} = \frac{(1-f)}{n} N^2 \{ S_o^2 - PQ\bar{W}^2 \}$$

por outro lado, de (8.9)

$$V(\hat{w}) = \frac{(1-f)}{n} N^2 S_o^2$$

comparando os dois últimos resultados teremos:

$$\frac{V(\tilde{w})}{V(\hat{w})} = \frac{S_o^2 - PQ\bar{W}^2}{S_o^2} = 1 - PQ\,\frac{\bar{W}^2}{S_o^2}$$

porém, lembrando que

$$\bar{W} = \frac{N}{M}\,\bar{W}_o = \frac{\bar{W}_o}{P} \quad \text{e fazendo } C_o = \frac{S_o}{\bar{W}_o}, \text{ vem}$$

$$\frac{V(\tilde{w})}{V(\hat{w})} = 1 - \frac{Q}{PC_o^2}$$

onde o fator $\frac{Q}{PC_o^2}$ indica o quanto decresce a variância quando se restringe a pesquisa apenas à parte não nula da população. Notamos que êsse decréscimo é maior, quando a proporção Q de nulos é grande. Ou quando o coeficiente de variação C_o é pequeno.

Evidentemente êste lucro na variância deve ser balanceado com o custo necessário para a eliminação, *a priori*, dos elementos nulos da população.

CAPÍTULO II

1. INTRODUÇÃO

Nêste capítulo iremos tratar da amostragem estratificada. Suponhamos nossa população N, dividida em k estratos mùtuamente exclusivos N_1, N_2, N_3, ..., N_k, contendo cada estrato respectivamente .... N_1, N_2, N_3, ..., N_k elementos. Logo o número total N de elementos da população é:

$$N = \sum_{i=1}^{k} N_i \qquad (1.1)$$

A subpopulação M interseccionará os estratos, em k partes mùtuamente esclusivas, M_1, M_2, M_3, ... M_k, cada uma contendo M_1, M_2, M_3, ..., M_k elementos. O número total M de elementos da subpopulação é dado por

$$M = \sum_{i=1}^{k} M_i \qquad (1.2)$$

Os N valôres da variável Y, serão indicados por dois índices. Assim Y_{ij} estará indicando a observação feita na j-ésima unidade do i-ésimo estrato. De acôrdo com o exposto acima Y_{ij} os índices terão os seguintes domínios: $i = 1, 2, \ldots k$ e $j = 1, 2, \ldots N_i$.

Como foi introduzido na primeira parte dêste trabalho, vamos definir a variável auxiliar:

$$W_{ij} = \begin{cases} Y_{ij} & \text{se} \quad Y_{ij} \in M \\ 0 & \text{se} \quad Y_{ij} \notin M \end{cases} \qquad (1.3)$$

O nosso problema consiste em estimar para a subpopulação, a média $\bar{W}$ ou o total W. Com :

$$\bar{W} = \frac{1}{M} \sum_{i=1}^{k} \sum_{j=1}^{N_i} W_{ij} = \frac{1}{M} \sum_{i=1}^{k} N_i \bar{W}_{io} = \frac{1}{M} \sum_{i=1}^{k} M_i \bar{W}_i \qquad (1.4)$$

e

$$W = \sum_{i=1}^{k} \sum_{j=1}^{N_i} W_{ij} = \sum_{i=1}^{k} M_i \bar{W}_i = \sum_{i=1}^{k} N_i \bar{W}_{io} \qquad (1.5)$$

onde

$$\bar{W}_i = \frac{1}{M_i} \sum_{j=1}^{N_i} W_{ij} \quad (1.6) \quad \text{e} \quad \bar{W}_{io} = \frac{1}{N_i} \sum_{j=1}^{N_i} W_{ij} \qquad (1.7)$$

Retiremos agora dentro de cada estrato, uma amostra casual simples, contendo n_1, n_2, ..., n_k elementos, respectivamente de cada

estrato. O número total n da amostra será:

$$n = \sum_{i=1}^{k} n_i \tag{1.8}$$

Essa amostra conterá m_1 elementos de M_1, m_2 elementos de M_2, etc. Dêsse modo o número m de elementos da subpopulação, que fazem parte da amostra é

$$m = \sum_{i=1}^{k} m_i \tag{1.9}$$

Através das mesmas considerações feitas na primeira parte do trabalho, notamos que fixado o tamanho n da amostra, o tamanho m é uma variável aleatória, bem como serão $m_1, m_2, \ldots, m_k$.

Ainda de acôrdo com a nomenclatura usada na primeira parte, onde as letras minúsculas eram usadas na representação da amostra, teremos:

$$\bar{w} = \frac{1}{m} \sum_{i=1}^{k} \sum_{j=1}^{n_i} w_{ij} = \frac{1}{m} \sum_{i=1}^{k} m_i \bar{w}_i = \frac{1}{m} \sum_{i=1}^{k} n_i \bar{w}_{io} \tag{1.10}$$

onde

$$\bar{w}_i = \frac{1}{m_i} \sum_{j=1}^{n_i} w_{ij} \qquad \text{e} \qquad \bar{w}_{io} = \frac{1}{n_i} \sum_{j=1}^{n_i} w_{ij} \tag{1.11}$$

ou para indicar o total da amostra:

$$w = \sum_{i=1}^{k} \sum_{j=1}^{n_i} w_{ij} = \sum_{i=1}^{k} m_i \bar{w}_i = \sum_{i=1}^{k} n_i \bar{w}_{io} \tag{1.12}$$

Usaremos ainda nêste capítulo, as notações:
a variável auxiliar:

$$X_{ij} = \begin{cases} 1 & \text{se } Y_{ij} \in M \\ 0 & \text{se } Y_{ij} \notin M \end{cases} \tag{1.13}$$

logo

$$M = \sum_{i=1}^{k} \sum_{j=1}^{N_i} X_{ij} = \sum_{i=1}^{k} N_i \bar{X}_i = \sum_{i=1}^{k} M_i$$

$$m = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} = \sum_{i=1}^{k} n_i \bar{x}_i = \sum_{i=1}^{k} m_i \tag{1.14}$$

as proporções :

$$P_i = \frac{M_i}{N_i} \quad \text{e} \quad Q_i = 1 - P_i = 1 - \frac{M_i}{N_i} \tag{1.15}$$

então $\bar{X}_i = P_i$

as frações amostrais

$$f_i = \frac{n_i}{N_i} \qquad \text{e} \qquad f_i' = \frac{m_i}{M_i} \tag{1.16}$$

2, ESTIMATIVA DE $\bar{W}$ COM OS M_i CONHECIDOS

Quando conhecem-se os valôres de M_i, Cochran (pág.147) sugere um processo que é a extensão das suposições feitas na secção I.4. Usando como estimador $\bar{W}$, $\bar{w}'$ definido do seguinte modo:

$$\bar{w}' = \frac{1}{M} \sum_{i=1}^{k} M_i \bar{w}_i \tag{2.1}$$

onde de acôrdo com as conclusões daquela secção, $\bar{w}_i$ é um estimador não viesado de $\bar{W}_i$, logo

$$E(\bar{w}') = \frac{1}{M} \sum_{i=1}^{k} M_i \, E(\bar{w}_i) = \frac{1}{M} \sum_{i=1}^{k} M_i \bar{W}_i = \bar{W} \tag{2.2}$$

e ainda de acôrdo com (I.4.3)

$$V(\bar{w}_i) = (1-f_i') \frac{S_{wi}^2}{m_i} \tag{2.3}$$

onde S_{wi}^2 indica a variância da variável W no i-ésimo estrato,

$$S_{wi}^2 = \frac{1}{M_i - 1} \sum_{j}^{M_i} (W_{ij} - \bar{W}_i)^2 \tag{2.4}$$

onde $\sum_{j}^{M_i}$ indica a somatória estendida apenas aos valôres não nulos de W_{ij}, no i-ésimo estrato.

Usando (2.3) a variância de $\bar{w}'$ reduz-se à

$$V(\bar{w}') = \frac{1}{M^2} \sum_{i=1}^{k} M_i^2 (1-f_i') \frac{S_{wi}^2}{m_i} \tag{2.5}$$

No caso particular de partilha proporcional ao tamanho dos estratos, e de variância constante em cada estrato, (2.5) se reduz a:

$$V(\bar{w}') = \frac{1}{Mf'} (1-f') S_w^2 = (1-f') \frac{S_w^2}{E(m)} \tag{2.6}$$

3. $\bar{w}$ COMO ESTIMADOR RAZÃO

Como desejamos estimar a média $\bar{W}$ da população, definida em

(1.7), uma sugestão é determinar um estimador não viesado para o numerador e outro para o denominador. Assim um estimador não viesado para:

$$W = \sum_{i=1}^{k} \sum_{j=1}^{N_i} W_{ij} = \sum_{i=1}^{k} N_i \bar{W}_{io} \qquad (3.1)$$

é

$$w = \sum_{i=1}^{k} N_i \bar{w}_{io} = \sum_{i=1}^{k} \frac{N_i}{n_i} \sum_{j=1}^{n_i} w_{ij} \qquad (3.2)$$

E um estimador não viesado para:

$$M = \sum_{i=1}^{k} \sum_{j=1}^{N_i} W_{ij} = \sum_{i=1}^{k} M_i = \sum_{i=1}^{k} N_i \bar{X}_i \qquad (3.3)$$

é

$$x = \sum_{i=1}^{k} N_i \bar{x}_i \qquad (3.4)$$

onde $\bar{x}_i = \frac{m_i}{n_i}$, mas conforme vimos na secção I.4

$$E(\bar{x}_i) = \frac{M_i}{N_i} \qquad (3.5)$$

então, calculando a esperança de x, temos imediatamente que

$$E(x) = M \qquad (3.6)$$

Vamos definir como estimador de $\bar{W}$, o quociente das expressões obtidas em (3.2) e (3.4), assim:

$$\bar{w}_e = \frac{\Sigma N_i \bar{w}_{io}}{\Sigma N_i \bar{x}_i} = \frac{\frac{1}{N} \Sigma N_i \bar{w}_{io}}{\frac{1}{N} \Sigma N_i \bar{x}_i} = \frac{\bar{w}_{es}}{\bar{x}_{es}} \qquad (3.7)$$

onde

$$\bar{w}_{es} = \frac{1}{N} \sum N_i \bar{w}_{io} \qquad (3.8)$$

e

$$\bar{x}_{es} = \frac{1}{N} \sum N_i \bar{x}_i \qquad (3.9)$$

Êste estimador,que é o estimador razão combinado devido a Hansen, Hurwitz e Gurney, é proposto por Cochran (pág.148) e por Durbin (pág. 114), para estimar $\bar{W}$.

Fava (pág.29) calculando a esperança matemática dêsse estimador, no caso geral, chegou em primeira aproximação ao seguinte resultado:

$$E(\bar{w}_e) = \bar{W} + \varepsilon \qquad (3.10)$$

onde ε, o viés, é dado pela seguinte expressão:

$$\varepsilon = \frac{1}{\bar{X}^2} \sum_i \frac{N_i(N_i-n_i)}{N\,n_i} \{\bar{W}\, S_{xi}^2 - S_{wxi}\} \qquad (3.11)$$

particularizando para a nossa variável,

$$S_{xi}^2 = \frac{1}{N_i-1} \{ \sum_{j=1}^{N_i} (X_{ij}-\bar{X}_i)^2 \} = \frac{M_i}{N_i-1} \left(1 - \frac{M_i}{N_i}\right) \cong P_i Q_i \qquad (3.12)$$

e

$$S_{wxi} = \frac{1}{N_i-1} \sum_{j=1}^{N_i} (W_{ij}-\bar{W}_{io})\ (X_{ij}-\bar{X}) =$$

$$= \frac{N_i \bar{W}_{io}}{N_i-1} \left(1 - \frac{M_i}{N_i}\right) \cong P_i Q_i \bar{W}_i \qquad (3.13)$$

substituindo êstes dois resultados em (3.11), vem :

$$\varepsilon = \frac{N^2}{M^2} \sum_i \frac{N_i(N_i-n_i)}{Nn_i} P_i Q_i (\bar{W}-\bar{W}_i) = \frac{N}{M^2} \sum_i \frac{N_i^2}{n_i} (1 - f_i)\ P_i Q_i\ (\bar{W}-\bar{W}_i) \qquad (3.14)$$

Caso a estratificação tenha sido feita com partilha proporcional, isto é,

$$f_i = \frac{n_i}{N_i} = \frac{n}{N} \qquad \text{para} \quad i = 1, 2, \ldots, k$$

então

$$\varepsilon = \frac{N^2}{M^2} \frac{(1 - f)}{n} \sum_i N_i P_i Q_i\ (\bar{W}-\bar{W}_i) \qquad (3.15)$$

ou lembrando que $P_i\ Q_i \leq \frac{1}{4}$, sempre, vem:

$$\varepsilon \leq \frac{1}{4} \frac{N^2}{M^2} \frac{(1 - f)}{n} \sum_i N_i\ (\bar{W} - \bar{W}_i)$$

o que nos permite ver mais claramente, a magnitude do viés em função de seus fatôres componentes.

Se a amostragem é feita com partilha proporcional e as variâncias e covariâncias são constantes de extrato para extrato é fácil mostrar usando raciocinio visto no Capítulo I que o viés é zero.

4. CÁLCULO DA VARIÂNCIA DE $\bar{w}_i$.

Queremos determinar :

$$V(\bar{w}_e) = E(\bar{w}_e - \bar{W})^2$$

mas

$$\bar{w}_e - \bar{W} = \frac{\bar{w}_{es}}{\bar{x}_{es}} - \bar{W} = \frac{\bar{w}_{es} - \bar{W}\bar{x}_{es}}{\bar{x}_{es}}$$

substituamos a variável $\bar{x}_{es}$ pela sua esperança $\bar{X}$. Assim:

$$\bar{w}_e - \bar{W} = \frac{1}{\bar{X}} (\bar{w}_{es} - \bar{W}\bar{x}_{es}) = \frac{1}{\bar{X}} \bar{u}_{es}$$

onde $\bar{u}_{es} = \bar{w}_{es} - \bar{W}\bar{x}_{es}$

então
$$V(\bar{w}_e) = \frac{1}{\bar{X}^2} V(\bar{u}_{es}) \qquad (4.1)$$

e $\bar{u}_{es}$ representa a média de uma amostra da variável aleatória $W_{ij} - \bar{W}X_{ij}$, colhida através de amostragem estratificada. Assim:

$$V(\bar{u}_{es}) = \frac{1}{N^2} \sum_i \frac{N_i(N_i - n_i)}{n_i} S_{ui}^2 \qquad (4.2)$$

onde
$$S_{ui}^2 = \frac{1}{N_i - 1} \sum_j^{N_i} \left\{ (W_{ij} - \bar{W}X_{ij}) - (\bar{W}_{io} - \bar{W}\bar{X}_i) \right\}^2 =$$

$$= \frac{1}{N_i - 1} \sum_j^{N_i} \left\{ (W_{ij} - \bar{W}X_{ij})^2 - N_i (\bar{W}_{io} - \bar{W}\bar{X}_i)^2 \right\} =$$

$$= \frac{1}{N_i - 1} \sum_j^{M_i} \left\{ (W_{ij} - \bar{W}_i)^2 + M_i (\bar{W}_i - \bar{W})^2 - N_i P_i^2 (\bar{W}_i - \bar{W})^2 \right\} =$$

$$= \frac{1}{N_i - 1} \sum_j^{M_i} \left\{ (W_{ij} - \bar{W}_i)^2 + M_i Q_i (\bar{W}_i - \bar{W})^2 \right\} \qquad (4.3)$$

substituindo (4.3) em (4.2), e o resultado em (4.1). vem:

$$V(\bar{w}_e) = \frac{1}{M^2} \sum_i \frac{N_i^2 (1 - f_i)}{n_i (N_i - 1)} \left\{ \sum_j^{M_i} (W_{ij} - \bar{W}_i)^2 + M_i Q_i (\bar{W}_i - \bar{W})^2 \right\} \qquad (4.4)$$

que é a variância procurada. Cochran (pág. 149) sugere como estimador dessa variância a expressão:

$$v(\bar{w}_e) \cong \frac{1}{x^2} \sum_i \frac{N_i^2 (1 - f_i)}{n_i (n_i - 1)} \left\{ \sum_j^{m_i} (w_{ij} - \bar{w}_i)^2 + m_i \left(1 - \frac{m_i}{n_i}\right) (\bar{w}_i - \bar{w}_e)^2 \right\}$$

onde x é dado por (3.4):

$$x = \sum_i^k N_i \bar{x}_i = \sum_i^k \frac{N_i}{n_i} \; m_i$$

Durbin (pág. 115) procurou medir qual o aumento sofrido pela variância devido a aleatoriedade de m. Para isso comparou a variância do estimador $\bar{w}_e$ com $\bar{w}'$, obtido na secção 2. Mas para isso é necessário algumas aproximações. Suponhamos que a estratificação foi tomada, em ambos os casos, com partilha proporcional ao tamanho do estrato, e que $f' \doteq f$. Nestas condições, de (4.4), vem:

$$V(\bar{w}_e) = \frac{1}{M^2} \frac{(1-f)}{f} \left\{ \sum_i^k \sum_j^{M_i} (W_{ij} - \bar{W}_i)^2 + \sum_i M_i Q_i (\bar{W}_i - \bar{W})^2 \right\} \qquad (4.5)$$

e de (2.5)

$$V(\bar{w}') = \frac{1}{M^2} \frac{(1-f)}{f} \sum_i^k \sum_j^{M_i} (\bar{W}_{ij} - \bar{W}_i)^2 \qquad (4.6)$$

logo

$$\frac{V(\bar{w}_e)}{V(\bar{w}')} = 1 + \frac{\sum_i^k M_i Q_i (\bar{W}_i - \bar{W})^2}{\sum_i^k \sum_j^{M_i} (W_{ij} - \bar{W}_i)^2}$$

e o último quociente, é que mede o aumento da variância devido ao desconhecimento do tamanho dos estratos subpopulacionais.

5. ESTIMATIVA DO TOTAL W DA SUBPOPULAÇÃO

5.1 Usando o fator de crescimento $\frac{N_i}{n_i}$

Quando desconhecem-se os valôres M_i, podemos usar o fator de crescimento $\frac{N_i}{n_i}$, e aplicando as mesmas considerações da secção I.8.3, dentro de cada estrato, na estimativa do total $\bar{W}$ da subpopulação, podemos usar o estimador não viesado

$$\hat{\bar{w}} = \sum_i^k \frac{N_i}{n_i} \sum_j^{n_i} w_{ij} \qquad (5.1)$$

e de acôrdo com o que vimos naquela secção, a variância dêsse estimador será dado por:

$$V(\hat{\bar{w}}) = \sum_i \frac{N_i^2 (1 - f_i)}{n_i} S_{wio}^2 \qquad (5.2)$$

com

$$S_{wio}^2 = \frac{1}{N_i - 1} \sum_j^{N_i} (W_{ij} - \bar{W}_{io})^2 = \frac{1}{N_i - 1} \left\{ (N_i - 1) S_{wi}^2 + M_i Q_i \bar{W}_i^2 \right\} \qquad (5.3)$$

ou seja,

$$V(\hat{\bar{w}}) = \sum_{i}^{k} \frac{N_i^2(1-f_i)}{n_i(N_i-1)} \{(M_i-1)\ S_{wi}^2 + M_i Q_i \bar{W}_i^2\} \qquad (5.4)$$

Êste estimador $\hat{\bar{w}}$, está em Cochran (pág.147), que apresenta como estimador da variância (5.2), a expressão:

$$v(\hat{\bar{w}}) = \sum_{i}^{k} \frac{N_i^2(1-f_i)}{n_i(n_i-1)} \left\{ \sum_{j}^{m_i} w_{ij}^2 - n_i \bar{w}_{io}^2 \right\} \qquad (5.5)$$

5.2 Usando o estimador razão separado

Para estimar o total da subpopulação quando conhecem-se os valôres M_i, podemos usar o estimador

$$w_s = \sum_{i}^{k} M_i \bar{w}_i = \sum_{i}^{k} M_i \frac{\bar{w}_{io}}{\bar{x}_i} \qquad (5.6)$$

que identificamos imediatamente como estimador razão separado (conf. Fava, pág.28). Vamos determinar a esperança e a variância dêsse estimador, a partir dos resultados chegados por Fava (pág.30), para estimador razão separado no caso geral. Assim, em primeira aproximação

$$E(w_s) = W + \varepsilon_s \qquad (5.7)$$

onde

$$\varepsilon_s = \sum_{i}^{k} \varepsilon_i = \sum_{i}^{k} \frac{N_i-n_i}{n_i\ \bar{X}_i} (R_i S_{xi}^2 - S_{wxi}) \qquad (5.8)$$

com

$$R_i = \frac{\bar{W}_{io}}{\bar{X}_i}$$

$$S_{xi}^2 = \frac{1}{N_i-1} \sum_{j}^{N_i} (X_{ij}-\bar{X}_i)^2 = \frac{M_i}{N_i-1} Q_i$$

$$S_{wi} = \frac{1}{N_i-1} \sum_{j}^{N_i} (W_{ij}-\bar{W}_{io})(X_{ij}-\bar{X}_i) \qquad (5.9)$$

Mas a expressão entre parêntesis da equação (5.8), é nula, pois dentro de cada estrato, a reta de regressão de W em função de X passa pela origem, e estamos na situação vista na secção I.5. Ou seja, w_s é em primeira aproximação, um estimador não viesado de W.

Para se ter uma idéia da magnitude do verdadeiro viés, podemos fazer as seguintes suposições:

- $\varepsilon_i \simeq \varepsilon$, para todo i = 1, 2, ..., k, isto é, temos aproximadamente o mesmo viés de estrato para estrato;
- A variância de W dentro de cada estrato, é a mesma;
- C.V.($\bar{x}_i$) ≃ C.V.($\bar{x}$), isto é, também o coeficiente de variação das médias de x em cada estrato são aproximadamente iguais.

Nessas condições teremos

$$\frac{|\varepsilon_s|}{\sqrt{V(w_s)}} \leq \frac{\sqrt{k}\, Q}{n} \qquad (5.10)$$

sendo esta uma expressão para indicar a magnitude do verdadeiro viés.

A partir da expressão 2.3.2 de Fava (Pág.31), determinamos a variância de w_s

$$V(w_s) = \sum_i^k \frac{N_i^2 (1-f_i)}{n_i (N_i-1)} (M_i-1) S_{wi}^2 \qquad (5.11)$$

Se compararmos êste resultado com (5.4), teremos uma expressão para indicar em quanto diminui a variância pelo fato de conhecermos os valôres M_i.

-o-

CAPÍTULO III

AMOSTRAGEM EM DOIS ESTÁGIOS

Vamos desenvolver aqui a mesma técnica dada por Durbin (pág.117). Consideremos inicialmente a amostragem em dois estágios, na qual k unidades são selecionadas, com reposição, na primeira fase de uma população de K unidades, com probabilidade de seleção proporcional aos números $N_1, N_2, \ldots, N_K$, que são os tamanhos das unidades amostrais da primeira fase. Na segunda fase é selecionada, em cada unidade da primeira fase, um número fixo n de elementos. Do mesmo modo como foi apresentado em II.1, queremos estimar parâmetros de uma subpopulação M. Por exemplo a média da subpopulação

$$\bar{W} = \frac{1}{M} \sum_{i=1}^{K} \sum_{j=1}^{N_i} W_{ij} \qquad (1.1)$$

e pode ser estimado por:

$$\bar{w}_1 = \frac{1}{m} \sum_{i=1}^{k} \sum_{j=1}^{n} w_{ij} \qquad (1.2)$$

Numa primeira ordem de aproximação, podemos dizer que:

$$E(\bar{w}_1) = \frac{E\left(\sum_i^k \sum_j^n w_{ij}\right)}{E(m)} = \frac{\frac{kn}{N} \sum_i^K \sum_j^{N_i} W_{ij}}{\frac{kn}{N} M} = \bar{W} \qquad (1.3)$$

ou seja, é um estimador aproximadamente não viesado. Para calcular a variância usaremos processo semelhante ao desenvolvido na secção II.4, assim

$$\bar{w}_1 - \bar{W} = \frac{\sum\sum w_{ij}}{\sum\sum x_{ij}} - \bar{W} = \frac{\sum\sum (w_{ij} - \bar{W}x_{ij})}{m} =$$

$$= \frac{k}{m} \frac{1}{k} \sum\sum u_{ij} \qquad (1.4)$$

onde $u_{ij} = w_{ij} - \bar{W}x_{ij}$. E a variância será dada por

$$v(\bar{w}_1) = \left(\frac{k}{m}\right)^2 v(\bar{u}) \qquad (1.5)$$

onde $\bar{u}$ é a média aritmética simples de k valôres da forma:

$u_i = \sum_{i=1}^{n} u_{ij} = w_i - \bar{W}m_i$, e de acôrdo com a teoria da amostragem:

$$v(\bar{u}) = \frac{1}{k}\,\frac{1}{k-1}\sum_{i}^{k}(u_i - \bar{u})^2 \qquad (1.6)$$

mas

$$\sum_{i}^{k}(u_i-\bar{u})^2 = \sum u_i^2 - k\bar{u}^2 = \sum (w_i-m_i\bar{W})^2 - \frac{m^2}{k}(\bar{w}_1-\bar{W})^2 \qquad (1.7)$$

Mas esta expressão contém o valor desconhecido $\bar{W}$, que não pode ser simplesmente substituido por $\bar{w}_1$ que é viesado. Entretanto, se os valôres m_i forem iguais entre si, e portanto iguais a $\frac{m}{k}$, teremos

$$\sum_{i}^{k}(w_i - \frac{m}{k}\bar{W})^2 = \sum (w_i - \frac{m}{k}\bar{w}_1)^2 + \frac{m^2}{k}(\bar{w}_1 - \bar{W})^2$$

então, é razoável supor válida a relação:

$$\sum_{i}^{k}(w_i-m_i\bar{W})^2 \cong \sum (w_i-m_i\bar{w}_1)^2 + \frac{m^2}{k}(\bar{w}_1-\bar{W})^2 \qquad (1.8)$$

substituindo este último resultado em (1.7), vem

$$\sum_{i}^{k}(u_i - \bar{u})^2 = \sum (w_i - m_i\,\bar{w}_i)^2$$

ou seja a variância estimada em (1.5) reduz-se a

$$v(\bar{w}_1) = \frac{k}{m^2(k-1)}\sum_{i}^{k}(w_i - m_i\,\bar{w}_1)^2 \qquad (1.9)$$

que é o estimador da variância procurada.

.-oOo-.

BIBLIOGRAFIA:

Cochran, W.G - SAMPLING TECHNIQUES - second edition - John Wiley & Sons - 1965.

Durbin, J. SAMPLING THOERY FOR ESTIMATES BASED ON FEWER INDIVIDUAL THAN THE MEMBER SELECTED - Bulletin of International Statistical Institute, Vol.36, pág. 113-119 - 1958.

FAVA, L. - CONTRIBUIÇÃO PARA O ESTUDO DA ESTIMATIVA RAZÃO - Tese apresentada para Concurso de livre docência - F.F.C.L. - U.S.P. - 1959.

Feller, W. - AN INTRODUCTION TO PROBABILITY THEORY AND APPLICATIONS John Wiley & Sons - Vol. 1 - 1950.

Hansen, M.H. - SAMPLE SURVEY, METHODS AND THEORY (Vol. I e II) John
Hurwitz,W.N. Wiley & Sons , 1953.
Madow, W.G.

Kish, L. - SURVEY SAMPLING - John Wilwy & Sons, 1967.

Sukhatme, P.V. - SAMPLING THEORY OF SURVEY WITH APPLICATIONS - The Indian Society of Agricultural Statistics, New Delhi, 1954.

YATES, F. - SAMPLING METHODS FOR CENSUSES AND SURVEYS - Charles Griffin and Co., London, 1953.

-0-