

## MÉTODOS DE AMOSTRAGEM PARA ESTIMAÇÃO DA COBERTURA VACINAL

Eunice Pinho de Castro Silva\*

SILVA, E.P. de C. Métodos de amostragem para estimação da cobertura vacinal. Rev.Saúde públ., S.Paulo, 20: 377-84, 1986.

**RESUMO:** Apresentam-se sucintamente dois métodos de amostragem que visam a seleção de uma amostra de crianças de fixada faixa de idade, residentes em determinada área geográfica de interesse, para estimação de cobertura vacinal: o método de R.H. Henderson e T. Sundaresan, o método do Departamento de Epidemiologia e de Métodos Quantitativos em Saúde da Escola Nacional de Saúde Pública. Um terceiro método de amostragem é proposto. O primeiro método apresentado (Henderson e Sundaresan) é empregado no Programa Ampliado de Imunização e nesse programa é considerado eficiente, simples e não dispendioso. O segundo e o novo método, que apresentam modificações do primeiro, constituindo alternativas deste, visam diminuir o erro quadrático médio nas estimativas, conquanto sejam menos simples e mais dispendiosos que aquele. Visto que, na estimação da cobertura vacinal, o estimador empregado pressupõe amostra autoponderada, a preocupação maior do método aqui proposto foi a de proporcionar uma amostragem segundo a qual se tenha equiprobabilidade de seleção para qualquer criança do grupo etário estudado residente na área de interesse, independentemente de qualquer condição.

**UNITERMOS:** Vacinação. Técnica de estimação. Amostragem. Probabilidade.

São, aqui apresentados métodos que visam a seleção de uma amostra de crianças de fixada faixa etária, residentes em determinada área geográfica, cujos dados proporcionarão estimativa da cobertura vacinal nessa área.

Inicialmente é apresentado o método de Henderson e Sundaresan, empregado na avaliação da cobertura vacinal no Programa Ampliado de Imunização<sup>3</sup>. O segundo, também aplicado nesse Programa, apresenta, na segunda etapa, uma modificação do primeiro<sup>5</sup>. Um terceiro método, com o mesmo objetivo, é aqui proposto, e constitui, assim como o segundo, uma alternativa do método inicial.

### 1. MÉTODOS UTILIZADOS

#### 1.1 Descrição Sucinta dos Métodos Utilizados

##### 1.1.1. Método de Henderson e Sundaresan

O Programa Ampliado de Imunização, que tem como especial interesse a eficácia dos serviços de imunização para crianças\*\* em países em desenvolvimento, procura métodos de implementação e avaliação de programas que sejam eficientes, simples e não dispendiosos. É citado como exemplo de um tal método de avaliação, satisfazendo essas condições, o método de Henderson e Sundaresan<sup>3</sup>. Este processo, ao qual nos referiremos chamando-o "método inicial", consiste em<sup>3,6</sup>:

a) identificação da área de interesse;

b) divisão da área de interesse em áreas parciais (unidades primárias de amostragem) para cada uma das quais se tem a respectiva população;

c) seleção casual de 30 dessas áreas parciais com probabilidade proporcional às respectivas populações;

d) seleção casual de um "ponto de partida" (domicílio inicial) em cada área parcial selecionada e, deste modo, todas as casas da área têm a mesma probabilidade de ser a primeira casa a ser visitada nessa área parcial<sup>6</sup>; o método usado para selecionar a primeira casa varia de acordo com a densidade populacional (zona urbana ou rural) e com a informação disponível para o pesquisador<sup>6</sup>;

e) seleção de 7 crianças em cada uma das 30 áreas parciais

— em cada área parcial selecionada os domicílios são considerados em fixada ordem (ordem de proximidade);

— a seleção, numa área parcial, começa no domicílio inicial e continua pelos domicílios seguintes até ser obtido um total de 7 crianças. Todas as crianças, residindo no último domicílio, são incluídas na amostra, mesmo que isto signifique incluir 8 ou mais crianças no conglomerado, em lugar do número requerido de 7 crianças.

\* Do Departamento de Epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo – Av. Dr. Arnaldo, 715 – 01255 – São Paulo, SP - Brasil.

\*\* NOTA: Sempre que constar o termo "criança" deverá ser entendido que se trata de criança do grupo etário estudado.

### 1.1.2. Primeiro método alternativo

Este método, proposto por Szwarcwald e Valente<sup>5</sup>, Departamento de Epidemiologia e Métodos Quantitativos em Saúde da Escola Nacional de Saúde Pública, adota as mesmas fases iniciais do primeiro método porém, antes da quarta fase, requer um pequeno censo em cada uma das 30 áreas parciais selecionadas, para identificação de todos os domicílios aí localizados, nos quais residam crianças. Dos domicílios, então arrolados, são selecionados sete deles com aplicação de amostragem sistemática.

### 1.2 Probabilidade de Seleção das Crianças nos Métodos Utilizados.

Notação:

$P_i$  população da área parcial  $i$  em determinado período de tempo  $T$  (o mais próximo da época do estudo e cujos dados populacionais são mais fidedignos).

$P_o = \sum_i^M P_i$  população total da área de interesse no período  $T$

$M$  número de áreas parciais (Unidades Primárias de Amostragem) nas quais a área de interesse foi dividida

$D_i$  número de domicílios na área parcial (UPA)  $i$  no período  $T$

$D_o = \sum_i^M D_i$  total de domicílios na área de interesse, no período  $T$

$D_i^1$  número atual de domicílios na área parcial (UPA)  $i$ , isto é, número de domicílios no período da pesquisa

$D_{ci}^1$  número atual de domicílios com criança, isto é, número de domicílios com criança no período de pesquisa

#### 1.2.1. Probabilidade de seleção no primeiro método

##### 1ª etapa

$\frac{30 P_i}{P_o}$  número esperado de vezes que a área parcial  $i$  entra na amostra de 30 áreas parciais

##### 2ª etapa

$\frac{1}{D_i^1}$  probabilidade que cada domicílio da área parcial  $i$  tem de ser domicílio inicial, dado que a área parcial  $i$  foi selecionada na primeira etapa.

Conseqüentemente, a probabilidade que cada domicílio da área parcial  $i$  tem de ser domicílio inicial é dada pela expressão (1)<sup>4</sup>

$$\frac{30 P_i}{P_o} \cdot \frac{1}{D_i^1} \quad (1)$$

A expressão (1) terá sempre o mesmo valor  $f$ , isto é, a amostragem de domicílios será equiprobabilística, se a razão  $P_i / D_i^1$  tiver o mesmo valor para todas as  $M$  áreas parciais, ou seja, se houver proporcionalidade direta entre a população da área parcial (referente a um período  $T$  de tempo) e o número de domicílios existentes nessa área no período do levantamento, podendo-se substituir  $P_i$  por  $c D_i^1$  onde  $c$  é uma constante. Nessas condições a expressão (1) ficará

$$\frac{30 c D_i^1}{P_o} \cdot \frac{1}{D_i^1} = \frac{30 c}{P_o} = f \quad (2)$$

Neste primeiro método, uma vez selecionado o domicílio inicial, a partir dele, seguindo determinada ordem, é tomado um conjunto de domicílios com crianças. A esse conjunto pertencerá o domicílio inicial se nele residir pelo menos uma criança. Se um domicílio com criança é selecionado, todas as crianças aí residentes são consideradas na amostra. Conseqüentemente, a probabilidade que uma criança, residente na área de interesse, tem de pertencer à amostra, é dada pela probabilidade que o domicílio, onde reside, seja selecionado.

Admitindo que qualquer domicílio da área estudada tenha igual probabilidade de ser um domicílio inicial, ou seja, sendo o valor da expressão (1) o mesmo para todas as  $M$  áreas parciais, ainda assim a probabilidade que um domicílio, com criança, tem de pertencer à amostra, vai depender do número de domicílios sem crianças intercalados entre os que têm crianças, bem como do número de crianças residentes nestes últimos.

Considerando a expressão (1) e o que consta no parágrafo anterior, concluímos que a probabilidade que uma criança, residente na área de interesse, tem de pertencer à amostra, vai depender de três fatores:

- $P_i / D_i$
- número de domicílios sem criança intercalados entre os com criança, que precedem aquele em que reside a criança
- número de crianças em um mesmo domicílio, nos domicílios com criança que precedem o desta

### 1.2.2. Probabilidade de seleção no segundo método

A modificação que o segundo método introduz no primeiro faz com que a probabilidade que cada criança residente na área de interesse, tem de ser selecionada, seja dada pela expressão (3)<sup>4</sup>

$$\frac{30 P_i}{P_0} = \frac{7}{D_{ci}} \quad (3)$$

Teremos ou não equiprobabilidade na seleção de crianças, conforme as razões  $P_i / D_{ci}$  (para  $i = 1, 2, \dots, M$ ) sejam ou não todas iguais entre si.

## 2. MÉTODO PROPOSTO

Este método que está apresentado levou em conta que o estimador empregado para estimar a cobertura vacinal, que é dado pela expressão (4), pressupõe que a amostragem tenha sido equiprobabilística de crianças<sup>2</sup>.

P cobertura vacinal

$\hat{P}$  estimador do valor populacional P

$$\hat{P} = \frac{\sum_j y_j}{\sum_j x_j} \quad (4)$$

$x_j$  número de crianças selecionadas na área parcial j

$y_j$  número de crianças vacinadas dentre as  $x_j$  selecionadas dessa área parcial j

Analogamente, este método levou em consideração que a fórmula do estimador da variância de  $\hat{P}$  pressupõe que todas as crianças tenham tido igual probabilidade de pertencer à amostra. Houve, portanto, ao propor este método, a preocupação de que fosse um método de amostragem equiprobabilística. Este método alternativo do método inicial contém as operações a seguir descritas:

1º) Determina-se a fração global de amostragem

$$f = \frac{n}{C} \quad (5)$$

onde

$$n = 210 \div a\%$$

210 tamanho desejado para a amostra de crianças

a% taxa de resposta em levantamento análogo

$\hat{C}$  estimativa de C

C total de crianças da faixa etária estudada durante o período em que se realiza o levantamento

A fração f será tomada como fração global de amostragem a ser aplicada aos domicílios com crianças residentes na área de interesse, os quais devem ser selecionados em duas etapas.

2º) Toma-se como unidade de primeira etapa (ou Unidade Primária de Amostragem UPA) o setor censitário ou conjunto de setores censitários.\*

A população  $P_i$ , que engloba todos os grupos etários, dada pelo Boletim Resumo, referente à i-ésima UPA, deverá ser tal que a estimativa de crianças nela residentes seja o dobro do número de crianças que se pretende sortear em cada UPA, ou seja:

$$P_i \text{ t\%} > 14 \quad (6)$$

onde t% é a proporção que o grupo etário das crianças estudadas representa na população geral.

Se um setor censitário não satisfizer essa condição, ele deverá ser reunido a um ou mais setores a fim de que o conjunto formado atenda àquelas condições e este conjunto constituirá uma UPA.

3º) Etapas do Levantamento<sup>4</sup>

**Primeira etapa**

Selecionam-se 30 UPA's com probabilidade proporcional ao "tamanho estimado" da UPA e que consiste no número de domicílios  $D_i$  dessa UPA, dado pelo "Boletim Resumo" (Veja Apêndice)

$$\frac{30 D_i}{D_0} \quad (7)$$

$D_i$  número de domicílios \*\* da UPA i

$$D_0 = \sum_i^M D_i \quad \text{total de domicílios na área de interesse}$$

M número de UPA's formadas com os setores censitários da área de interesse

\* A relação dos setores censitários, com indicação do número de domicílios  $D_i$  e respectiva população  $P_i$ , no período do Censo Demográfico, consta em "CD 1.10 - Boletim Resumo", elaborado pela Fundação IBGE - Instituto Brasileiro de Geografia e Estatística - Departamento de Censos.

\*\* Nota: Sempre que constar o termo "domicílio" deve-se entender que se trata de domicílio particular.

## Segunda etapa

Os domicílios de uma UPA selecionada na primeira etapa são considerados em determinada ordem e, implicitamente, os domicílios com crianças também ficarão ordenados. Exemplificando essa operação, se a UPA sorteada for em um setor censitário da zona urbana, para ordenar os domicílios, poderemos primeiramente ordenar os quarteirões e, em cada quarteirão, poderemos ordenar os domicílios a partir de fixada esquina e seguindo determinado sentido de percurso (sentido horário ou anti-horário), o qual, uma vez escolhido, deve ser o único adotado.

Aos domicílios com crianças existentes na UPA  $i$ , considerados na ordem estabelecida, aplica-se amostragem sistemática com fração de amostragem cujo denominador é  $D_i$  e cujo numerador é  $b$  tal que  $b/D_i$  multiplicado por (7) resulte o valor da fração global de amostragem  $f$ .

$$\frac{30 D_i}{D_0} \cdot \frac{b}{D_i} = f \quad (8)$$

ou seja

$$b = \frac{f D_0}{30}$$

Outro requisito que a UPA deve satisfazer é o de que  $D_i \geq b$ . Caso isso não ocorra, os setores censitários deverão ser reunidos de modo que, na UPA então formada, essa condição esteja satisfeita.

Para a aplicação da amostragem sistemática na UPA  $i$  selecionada, emprega-se o intervalo  $D_i/b$ . Prepara-se para cada UPA selecionada, a relação contendo o intervalo de amostragem sistemática, o começo casual sorteado e os "números selecionados". Dado que os domicílios com crianças na UPA estão considerados numa determinada ordem, aqueles que, pela respectiva numeração, correspondem aos "números selecionados", são os domicílios com criança que comporão a amostra.

## Observações

- 1ª - Se, na divisão de  $D_i$  por  $b$ , não resultar número inteiro exato, faz-se a aproximação até a quarta casa decimal, se até alguma anterior a esta, não tiver sido exata a divisão.
- 2ª - O começo casual deverá ter tantas casas decimais quantas forem as do intervalo.
- 3ª - Os "números selecionados" são obtidos tomando-se o começo casual, com as suas casas decimais (se for o caso) e a ele somando-se

sucessivamente o intervalo (também com as suas casas decimais se for o caso); na seqüência de números que vão sendo obtidos, cada qual com as suas casas decimais, toma-se apenas a parte inteira, porém não fazendo aproximação para o número inteiro mais próximo.

- 4ª - Se uma UPA, em decorrência de seu "tamanho estimado"  $D_i$ , for selecionada  $t$  vezes ( $t > 1$ ), na segunda etapa, aplica-se amostragem sistemática com fração de amostragem  $tb/D_i$ , portanto com intervalo igual a  $1/t$  daquele que seria aplicado se a UPA fosse selecionada apenas uma vez<sup>4</sup>.
- 5ª - Instruções para o percurso de cada UPA sorteada.  
No percurso de uma UPA sorteada, procede-se da forma a seguir descrita.
  - a) Na aplicação da amostragem sistemática, o percurso da UPA deve ser completo, de forma que a UPA toda seja percorrida.
  - b) Em todos os domicílios da UPA deve ser verificada a existência de criança.
  - c) Os domicílios fechados e portanto em relação aos quais não se sabe da existência ou não de crianças, devem ser registrados para formar um estrato a parte.
  - d) Os domicílios onde se constata que residem crianças devem ser registrados e numerados de forma que, com o percurso da UPA, vai sendo elaborada a lista dos domicílios com criança.
  - e) Na feitura dessa lista, quando é registrado um domicílio que, na numeração, recebe um número que é um dos "números selecionados", esse domicílio deve ser entrevistado pois é um domicílio da amostra.
  - f) Faz-se uma segunda tentativa para obter informação sobre a existência de crianças, nos domicílios fechados, procurando mudar o horário de averiguação. Desses domicílios, aqueles que tiveram crianças devem ir recebendo numeração em seqüência àquela dos domicílios com criança, não fechados na primeira tentativa, e assim ficam sujeitos à amostragem sistemática que estava sendo aplicada a essa UPA.  
Recomenda-se fazer mais duas tentativas com esse mesmo propósito.
  - g) Nos domicílios com crianças sorteadas, se a entrevista foi recusada ou houve algum impedimento para a sua realização, recomenda-se fazer mais até três retornos, procurando obter a informação.

6ª - Seja:

$$\bar{x} = \frac{\sum_{j=1}^{30} x_j}{30}$$

$x_j$  número de crianças selecionadas na UPA.  
 $j$  da amostra.

O coeficiente de variação de  $\bar{x}$  não deve superar 10%<sup>1</sup>, mas sendo inferior a 20% ainda é tolerável<sup>4</sup>. Se esse coeficiente não atender a esse requisito, o pesquisador deve estar alertado de que, se houve vício na estimativa-razão, este poderá ser não desprezível em relação ao erro padrão desse estimador e, além disso, uma das condições para a validade da fórmula que dá a variância do estimador, não estará sendo atendida.<sup>1,4</sup> Nessa hipótese, recomenda-se para levantamentos futuros, em áreas análogas, verificar qual o número de UPA's a serem selecionadas na primeira etapa, a fim de que aquele requisito seja atendido.

7ª - Se em uma UPA selecionada, o número de domicílios com crianças, existentes no período do levantamento  $D_{ij}$ , for menor que o intervalo  $D_i / b$ , pode ocorrer que o primeiro número selecionado seja maior que  $D_{ij}$ , resultando então que nenhum domicílio com criança seja selecionado nessa UPA. Com o intuito de que essa UPA, selecionada na primeira etapa, fique representada na amostra de domicílios com crianças, poderemos diminuir o intervalo da amostragem sistemática da segunda etapa, de forma que um ou mais daqueles domicílios sejam sorteados. Se entretanto o intervalo que era para ser aplicado, for substituído por um intervalo menor, a probabilidade de seleção dos domicílios com crianças dessa UPA ficará aumentada e este fato deve ser levado em conta ao ser feita a estimação da cobertura vacinal e, para tal, os valores encontrados na subamostra dessa UPA receberão um peso adequado, o qual será menor que o atribuído aos valores das subamostras das demais UPA's, para as quais não houve alteração do intervalo  $D_i / b$  correspondente. Observação análoga fazemos em relação à estimação da variância do estimador.

Todas as crianças do grupo etário estudado, residentes nos domicílios sorteados, devem ser consideradas da amostra.

De acordo com este método, todas as crianças da área de interesse têm igual probabilidade de pertencer à amostra, probabilidade esta dada pelo valor da fração  $f$ , e, portanto, a amostragem é equiprobabilística.

### 3. CONSIDERAÇÕES SOBRE OS MÉTODOS

#### 3.1 Considerações Sobre a Precisão e o Custo

O método inicial é de aplicação simples e requer menos dados que os demais. O segundo e o terceiro métodos precisam do percurso pela UPA toda e requerem a aplicação de amostragem sistemática na segunda etapa, fatos estes que os tornam menos simples e mais dispendiosos que o inicial. O terceiro método necessita do número de domicílios, dado este não requerido pelos demais métodos.

Admitindo a situação em que os dois primeiros métodos proporcionam amostras equiprobabilística e pressupondo que apliquem, assim como o método proposto, a expressão (4) para estimar a cobertura vacinal, passamos às considerações sobre precisão e custo.

Empregando na amostragem, conglomerados que consistem, cada qual, de unidades mais ou menos contíguas geograficamente, é comum existir uma correlação intraclasse positiva a qual geralmente é maior entre as unidades mais próximas fisicamente do que para unidades mais dispersas<sup>2</sup>. Desta forma, é de se esperar uma maior medida de homogeneidade quando o conglomerado é formado por 7 crianças tomadas nos domicílios em seqüência, a partir de um domicílio inicial, como no primeiro método, do que quando o conglomerado é formado por domicílios com crianças, os quais, por constituírem uma amostra sistemática<sup>1</sup>, estão dispersos por toda uma UPA como no segundo e terceiro métodos. Visto que na amostragem por conglomerado, para um mesmo número de elementos na amostra e para um mesmo tamanho médio de conglomerado, quanto maior a medida de homogeneidade dos conglomerados, maior é o efeito de planejamento e, portanto, menor a precisão<sup>2,4</sup>. Conclui-se que, com a aplicação do segundo e terceiro métodos, busca-se aumentar a precisão na estimação da cobertura vacinal.

#### 3.2 Considerações Referentes à Probabilidade de Seleção das Crianças.

O método inicial é probabilístico quando são conhecidos os fatores  $a$ ,  $b$ ,  $c$  citados em 1.2.1. Uma situação em que a amostragem nesse método será equiprobabilística, é aquela em que esses valores  $a$ ,  $b$  e  $c$  são os mesmos para todas as UPA's da área de interesse.

O segundo método é probabilístico, visto que o fator  $P_i/D_{ij}$  pode ser conhecido para todas as UPA's da área de interesse. A amostragem desse método será equiprobabilística se a razão  $P_i/D_{ij}$  for constante para todas as UPA's da área de interesse.

O terceiro método proporciona uma amostragem que é equiprobabilística, independentemente dos valores de  $D_i$  (número de domicílios na UPA) que serviram para seleção das UPA's na primeira etapa e na seleção dos domicílios na segunda etapa, figurando no numerador e no denominador da expressão (8). A apresentação do método proposto surgiu da preocupação em procurar alcançar equiprobabilidade na seleção de crianças para compor uma amostra com base na qual se fariam estimativas das coberturas vacinais. A equiprobabilidade é enfatizada em vista do fato de que o estimador<sup>4</sup> empregado na estimação da cobertura vacinal, dado pela expressão (4), pressupõe amostra autoponderada, ou seja, todas as crianças da área de interesse tendo igual probabilidade de pertencer à amostra<sup>2</sup>.

### 3.3 Consideração conclusiva

O método proposto requer mais dados que os dois primeiros, é mais dispendioso que o inicial e de aplicação menos simples que este, porém, é apresentado procurando aumentar a eficiência e, principalmente, visando obter uma amostra equiprobabilística de crianças, face ao estimador empregado na estimação da cobertura vacinal.

### APÊNDICE

Seleção de unidades primárias de amostragem com probabilidade proporcional ao número de domicílios.

Notação:

$D_i$  número de domicílios da unidade primária de amostragem (UPA)  $i$

$$D = \sum_i^M D_i$$

$M$  número de UPA's na área de interesse

$$\frac{D}{30} \text{ intervalo a ser aplicado para a seleção de UPA's}$$

### Observação:

Se  $D$  não for múltiplo de 30 e portanto o quociente  $D/30$  não for número inteiro, aproxima-se o mesmo, se for necessário, até a quarta casa decimal.

### Procedimento:

- 1º - Consideram-se as unidades primárias de amostragem em uma determinada ordem.
- 2º - Efetuam-se  $M$  operações de adição, cada qual correspondente a uma UPA. As parcelas da adição correspondente à  $i$ -ésima UPA são  $D_1, D_2, \dots, D_i$ .
- 3º - Determinam-se os  $M$  intervalos obtidos da forma que consta na relação apresentada, sendo que o extremo inferior do primeiro intervalo será igual a 1.

| UPA      | Valores acumulados de $D_i$ | Intervalo                                 |
|----------|-----------------------------|---|
| 1        | $D_1$                       | 1 ——— $D_1$                               |
| 2        | $D_1 + D_2$                 | $D_1 + 1$ ——— $D_1 + D_2$                 |
| $\vdots$ | $\vdots$                    | $\vdots$                                  |
| $i$      | $\sum_j^i D_j$              | $\sum_j^{i-1} D_j + 1$ ——— $\sum_j^i D_j$ |
| $\vdots$ | $\vdots$                    | $\vdots$                                  |
| $M$      | $\sum_j^M D_j$              | $\sum_j^{M-1} D_j + 1$ ——— $\sum_j^M D_j$ |

4º - Sorteia-se o começo casual  $r$ , sendo

$$10^{-u} \leq r \leq D/30$$

onde  $u$  é o número de casas decimais do intervalo  $D/30$  e  $r$  terá esse número de casas decimais.

5º - Ao número  $r$  soma-se sucessivas vezes o intervalo  $D/30$  até obter como resultado um número cuja parte inteira seja maior ou igual a  $D$ . Desta forma teremos obtido 30 números não inferiores a 1.

6º - Toma-se apenas a parte inteira dos 30 números selecionados (sem se aproximar obrigatoriamente para o número inteiro mais próximo) e assim se obtém os 30 "números selecionados".

7º - Procura-se a qual dos intervalos pertence cada um dos "números selecionados" e as UPA's correspondentes a estes intervalos serão as UPA's selecionadas.

Exemplo: De quatro UPA's cujo número de domicílios são respectivamente 50, 60, 40, 17 domicílios, selecionar três delas com probabilidade proporcional ao número de domicílios.

| UPA $i$ | $D_i$ | $\sum_j D_j$ | Intervalos  |
|---------|-------|--------------|-------------|
| I       | 50    | 50           | 1 ——— 50    |
| II      | 60    | 110          | 51 ——— 110  |
| III     | 40    | 150          | 111 ——— 150 |
| IV      | 17    | 167          | 151 ——— 167 |

$$D = 167$$

$$D/3 = 167/3 = 55,6667 \text{ (intervalo para seleção das UPA's)}$$

$$u \text{ (número de casas decimais)} = 4$$

$$10^{-4} \leq r \leq 55,6667$$

Suponhamos que  $r = 34,9899$ . Somando-se sucessivas vezes o intervalo teremos: 90,6566; 146,3233; 201,9900. Os números selecionados são: 34, 90, 146 e as UPA's selecionadas são: I, II, III.

Suponhamos que o número sorteado seja 009387 e portanto  $r = 0,9387$ . As somas sucessivas serão: 56,6054; 112,2721; 167,9388.

Os números selecionados serão 56, 112, 167 e, portanto, as UPA's selecionadas serão II, III e IV.

#### REFERÊNCIAS BIBLIOGRÁFICAS

1. COCHRAN, W.G. *Sampling techniques*. 2<sup>nd</sup> ed. New York, John Wiley & Sons, 1963.
2. HANSEN, M.H. et al. *Simple survey methods and theory*. New York, John Wiley & Sons, 1964. v. 1.
3. HENDERSON, R.H. & SUNDARENSAN, T. Cluster sampling to assess immunization a simplified sampling method. *Bull. Wld Hlth Org.*, 60: 253-60, 1982.
4. KISH, L. *Survey sampling*. New York, John Wiley & Sons, 1965.
5. SZWARCOWALD, C.L. & VALENTE, J.G. Avaliação da cobertura de vacinação em Teresina-Piauí (Brasil) 1983. *Cad. Saúde públ.*, Rio de Janeiro, 1 : 41-9, 1985.
6. WORLD HEALTH ORGANIZATION. Expanded Programme on Immunization. **Training for mid-level managers: evaluate vaccination coverage**. Geneva, s.d.

Recebido para publicação em 18/03/1986

Aprovado para publicação em 09/09/1986

SILVA, E.P. de C. [ Methods of sample selection for estimates of vaccination coverage ] . *Rev. Saúde públ.*, S. Paulo, 20 : 377-84, 1986.

**ABSTRACT:** It is presented two sampling methods for selecting a sample of children of a given age range, living in a geographical area of interest, to estimate vaccinal coverage: the R.H. Henderson and T. Sundaresan method, the Department of Epidemiology and Quantitative Methods of National Public Health School method. A third method of selection is then proposed. The geographical area of interest is divided into parts which are used as primary sampling units (PSU's). First method: 30 PSU's are selected with probability proportional to the population living in the PSU; a starting point ("household") is selected by random selection within each PSU selected: selection of 7 children from within each of the PSU's, begins with the starting household, and then continues to the next nearest household until a total of 7 individuals is obtained (all individuals of the appropriate age living in the last household falling within the sample are included). Second method: the first stage is like that of the first method; previously to the second stage there is a census in each PSU selected in order to prepare a list of all dwelling units where one or more children live; applying systematic selection, 7 dwelling units are selected from this list. Third method: 30 PSU's are selected with probability proportional to  $D_i$  (number of dwelling units in the PSU  $i$ ) given by the Census; the dwelling units in the selected PSU are considered in a fixed order; at the second stage, each selected PSU is exhaustively visited, according to the order of the dwelling units; during this visit, only the dwelling units where one or more children live are listed; to these dwellings a systematic selection is applied, the interval  $D_i/b$  of which is such that with these two stages there is an equal probability selection method (epsem); the over-all sampling fraction is  $n/\bar{C}$  with  $n$  given by  $210/a\%$  ( $a\%$  is the expected response rate) and  $\bar{C}$  is the estimate of the number of children living in the area of interest during the survey. The probability of selection of each individual of the appropriate age, in the first method, depends on the ratio  $P_i/D_i'$  ( $P_i$  is the population of the  $i^{\text{th}}$  PSU;  $D_i'$  is the number of dwelling units in this PSU, during the survey), it depends too on the number of dwelling units with no children living in them, situated between two dwellings with children, as well as on the number of children living in each dwelling unit. When the ratios  $P_i/D_i'$  are equal for all the PSU's, and the second item has the same values for the whole area of interest and this happens also with the third item, we shall then have a selection method of equal probability (epsem) to the first. The second method is a probability one, since  $P_i/D_{ci}'$  can be known for all the PSU's of the area of interest ( $D_{ci}'$  is the number of dwelling units with children living in, during the survey; this method will be an equal probability selection when that ratio is the same for all PSU's. The third method is an epsem, independently of the values of  $D_i$ . The first method is used by the Expanded Programme on Immunization and is considered effective, simple and inexpensive. The second and third present some changes and consist of alternative methods which attempt to diminish the mean square error of the estimator, although they are less simple and more expensive than the first one. The third requires more data than the second. Since the estimator used to estimate the vaccinal coverage presupposes self-weighting samples, the principal purpose of the third method was to provide a method which will be epsem, i.e., will give to every child of the age group studied, living in the area of interest, equal probability of selection.

**UNITERMS:** Vaccination. Estimation technics. Sampling studies. Probability.