

## AMOSTRAGEM ESTATÍSTICA

Maria Emilia Camargo

Departamento de Estatística. Centro de Ciências Naturais e Exatas.  
UFSM. Santa Maria, RS.

## RESUMO

Neste trabalho, serão abordados os principais aspectos da Teoria da Amostragem e explicitados os processos de cálculo utilizados na determinação da amostra nos casos de amostragem aleatória simples, estratificada com distribuição proporcional e ótima.

## SUMMARY

CAMARGO, M.E., 1987. Statistical Sample. *Ciência e Natura*, 9:7-19, 1987.

Principal aspects of sampling theory and calculating processes used in determining the size of a sample, categorized and group with proportional and optimum distribution in case of a random sample, are presented in this paper.

## INTRODUÇÃO

Os pesquisadores de todo o mundo, na realização de pesquisas científicas, em qualquer setor da atividade humana, utilizam as técnicas de amostragem no planejamento de seus trabalhos, não são pela impraticabilidade de poderem observar, numericamente, em sua totalidade determinada população em estudo, como devido ao aspecto econômico dessas investigações, conduzidos com um menor custo operacional, dentro de um menor tempo, além de possibilitar maior precisão nos respectivos resultados, ao contrário, do que ocorre com os trabalhos realizados pelo processo censitário (COCHRAN, 1965; CRUZ, 1978).

A técnica da amostragem, a despeito da sua larga utilização, ainda necessita de alguma didática mais adequada aos pesquisadores iniciantes. Assim, neste trabalho serão abordados os principais aspectos da Teoria da Amostragem e explicitada os processos de cálculo utilizados na determinação do tamanho da amostra nos casos de amostragem aleatória simples, estratificada com distribuição proporcional e com distribuição ótima. (GMURMAM, 1984; HOEL, 1962; DIXON & MASSEY, 1969; TUKEY; Y.W., 1977).

## DESENVOLVIMENTO

## ESPERANÇA MATEMÁTICA DE UMA FUNÇÃO LINEAR

Consideremos uma variável  $Z$  que seja função linear de  $n$  outras variáveis aleatórias  $X_1, X_2, \dots, X_n$ , isto é,

$$Z = a_1X_1 + a_2X_2 + \dots + a_nX_n$$

onde os "a<sub>i</sub>" são constantes quaisquer.

A sua esperança matemática é dada por

$$E(Z) = a_1 E(X_1) + a_2 E(X_2) + \dots + a_n E(X_n)$$

Representando a esperança matemática de E(Z) por  $\bar{Z}$  e a de E(X<sub>i</sub>) por  $\bar{X}$ , tem-se

$$Z = a_1 \bar{X}_1 + a_2 \bar{X}_2 + \dots + a_n \bar{X}_n$$

Substituindo esta última da 1ª, vem

$$Z - \bar{Z} = a_1 (X_1 - \bar{X}_1) + a_2 (X_2 - \bar{X}_2) + \dots + a_n (X_n - \bar{X}_n)$$

depois de se colocar os a<sub>i</sub> em evidência.

Fazendo,

$$(X_1 - \bar{X}_1) = W_1; (X_2 - \bar{X}_2) = W_2; \dots; (X_n - \bar{X}_n) = W_n$$

pode-se escrever:

$$Z - \bar{Z} = a_1 W_1 + a_2 W_2 + \dots + a_n W_n$$

Elevando-se ambos os membros ao quadrado,

$$(Z - \bar{Z})^2 = (a_1 W_1 + a_2 W_2 + \dots + a_n W_n)^2$$

e desenvolvendo o 2º membro, resulta:

$$(Z - \bar{Z})^2 = \sum_{i=1}^n a_i^2 W_i^2 + \sum_{i \neq j} a_i a_j W_i W_j$$

Substituindo W pelos seus valores:

$$(Z - \bar{Z})^2 = \sum_{i=1}^n a_i^2 (X_i - \bar{X}_i)^2 + \sum_{i \neq j} a_i a_j (X_i - \bar{X}_i) (X_j - \bar{X}_j)$$

Tomando a esperança matemática de cada membro:

$$E(Z - \bar{Z}) = \sum_{i=1}^n a_i E(X_i - \bar{X}_i) + \sum_{i \neq j} a_i a_j E(X_i - \bar{X}_i) (X_j - \bar{X}_j) \quad (1)$$

Sendo, por definição, o coeficiente de correlação igual a

$$\rho_{ij} = \frac{E[(X - \bar{X})(Y - \bar{Y})]}{\sigma_i \sigma_j}$$

tem-se que

$$\rho_{ij} \sigma_i \sigma_j = E[(X - \bar{X})(Y - \bar{Y})]$$

Substituindo o 2º membro na igualdade (1):

$$E[(Z - \bar{Z})^2] = \sum_{i=1}^n a_i^2 E(X_i - \bar{X}_i)^2 + \sum_{i \neq j} a_i a_j \rho_{ij} \sigma_i \sigma_j$$

ou, sabendo que

$$E[(Z - \bar{Z})^2] = \sigma_Z^2 \text{ e } E(X - \bar{X})^2 = \sigma_i^2$$

$$\sigma_{\omega}^2 = \sum_{i=1}^n a_i^2 \sigma^2 + \sum_{i \neq j} a_i a_j \rho_{ij} \sigma_i \sigma_j \quad (2)$$

Se as variáveis  $X_i$  e  $X_j$  são independentes duas a duas, o coeficiente de correlação será nulo e a fórmula se reduz a

$$\sigma_{\omega}^2 = \sum_{i=1}^n a_i^2 \quad (3)$$

As fórmulas (2) e (3) permitem calcular as variâncias das médias das amostras, tanto no caso de "extração sem reposição" quanto no de "extração com reposição".

#### MÉDIA E VARIÂNCIA DA DISTRIBUIÇÃO DE MÉDIAS

##### Extração com reposição

Seja uma população qualquer da qual se extrai uma amostra, com reposição do elemento extraído, de tamanho  $n$ . A média aritmética dessa amostra será dada por

$$\bar{X} = 1/n (x_1 + x_2 + \dots + x_n) = (1/n) \sum_{i=1}^n x_i$$

e representa uma estimativa da média da população.

Extraindo-se grande número de amostras obtêm-se valores diversos para  $x_i$ , nesse caso, a média aritmética  $\bar{X}$  pode ser considerada uma variável aleatória, uma vez que cada amostra extraída corresponde a um valor de  $\bar{X}$  com determinada probabilidade. Nessas condições, pode-se escrever:

$$\bar{X} = (1/n) (X_1 + X_2 + \dots + X_n) = (1/n) \sum_{i=1}^n (X_i) \quad (4)$$

sendo que os  $X_1, X_2, \dots, X_n$  representam agora variáveis aleatórias e não mais os valores obtidos numa amostra.

Tomando a esperança matemática de ambos os membros,

$$E(\bar{X}) = (1/n) E(X_1) + E(X_2) + \dots + E(X_n) \quad (5)$$

Assim, pelas próprias condições da amostragem  $E(X_i) = \bar{X}$  média da população para qualquer que seja o índice  $i$ .

Supondo-se que tivessem sido extraídas todas as amostras possíveis, com reposição, de um universo determinado, poder-se-ia, então, associar a cada extração uma certa variável: a extração de ordem 1, uma variável aleatória  $X_1$ ; a extração de ordem 2, uma variável  $X_2$ ; a de ordem  $i$ , uma variável  $X_i$ , e assim por diante.

A variável  $X_i$  assumirá todos os valores  $x_i$  da extração de ordem  $i$  em cada amostra, de modo que resulta  $E(X_i) = \bar{X}$  média da população para qualquer um dos valores  $i$ .

Analogamente se mostra que as variâncias das amostras são iguais às variâncias do universo, ou seja,  $\sigma^2 = \sigma^2$

Para calcular a variância da distribuição das médias

reconsidere-se a expressão dada em (4) e note-se que ela coincide com a variável  $Z$  em que todos os  $a_i$  são iguais, isto é,

$$a_1 = a_2 = \dots = a_n = 1/n \quad (6)$$

Mas, no caso de amostras com reposição, as variáveis  $X_1$  são independentes, o que equivale dizer que os coeficientes de correlação são nulos, prevalecendo, assim, a expressão dada em (3). Observe-se, porém, que todos os  $\sigma^2$  são iguais à variância da população. O termo  $a_i^2$ , em face de (6), é igual a  $1/n^2$ . Substituindo esses valores em (3), resulta:

$$\sigma_x^2 = \Sigma (1/n^2) \sigma^2$$

somatório de  $n$  termos constantes. Logo,

$$\sigma_x^2 = (1/n^2) n \sigma^2 = \sigma^2/n$$

expressão que fornece a variância da média da amostra em função da variância da população e do tamanho  $n$  da amostra. Como se vê, a variância da distribuição de médias é igual à variância da população dividida pelo número de elementos da amostra. A raiz quadrada a ambos os membros fornece o erro padrão da média.

#### *Extração sem reposição*

No caso de amostra sem reposição do elemento extraído, as extrações não são independentes, de modo que o coeficiente de correlação não é nulo, mas constante. Significa isto que se forem extraídas todas as amostras possíveis de uma população, sem reposição, e considerados os valores obtidos em todas as primeiras extrações como uma variável aleatória  $X_1$ ; os valores obtidos em todas as segundas extrações como uma variável aleatória  $X_2$ , etc. Os coeficientes de correlação entre  $X_1$  e  $X_2$ ,  $X_1$  e  $X_3$ , ...,  $X_1$  e  $X_n$ ,  $X_2$  e  $X_3$ ,  $X_2$  e  $X_4$ , ...,  $X_{n-1}$  e  $X_n$  serão iguais.

Fazendo na expressão geral (2)  $a_i = 1/n$  e  $a_j = 1/n$  e substituindo  $\rho$  e  $\sigma$  por  $\rho$ , visto que eles são constantes, e  $\rho$  por  $\sigma$ , resulta

$$\sigma_x^2 = \sum_{i=1}^n (1/n)^2 + \sigma^2 \sum_{i \neq j} (1/n) (1/n) \rho$$

isto é, todos os termos no interior de cada somatório são constantes. Sendo  $n$  o número de termos do 1º somatório e  $n(n-1)$  o do 2º uma vez que no caso de extrações sem reposição não figuram elementos repetidos, a expressão assume a forma:

$$\sigma_x^2 = (n/n^2)\sigma^2 + (\sigma^2/n^2) \rho \cdot n(n-1)$$

ou, simplificando,

$$\sigma_x^2 = (\sigma^2/n) [1 + \rho(n-1)] \quad (7)$$

A fim de determinar o valor de  $\rho$  supõe-se que  $n = N$ , ou seja, que são extraídas amostras compostas de todos os elementos da população. Nesse caso, em que a amostra esgota a população, a variância da média da amostra é nula, pois a média da amostra coincide com a da população, não havendo, por conseguinte, dispersão em torno dessa medida.

Substituindo  $n$  por  $N$  em (7) e  $\sigma_x^2$  por 0, vem

$$0 = \sigma^2/N[1 + \rho(N - 1)]$$

o que conduz, no caso de ser  $\sigma^2 \neq 0$ , a

$$0 = 1 + \rho(N - 1)$$

donde

$$\rho = -[1/(N-1)]$$

Substituindo em (7) o valor de  $\rho$ , resulta:

$$\sigma_x^2 = \sigma^2/n [1 - (1/(N-1))(n-1)] = \sigma^2/n [1 - (n-1)/(N-1)]$$

ou, simplificando

$$\sigma_x^2 = (\sigma^2/n) (N - n)/(N - 1) \quad (8)$$

expressão que fornece a variância da média da amostra, no caso de extração sem reposição, em função da variância, do tamanho da amostra e do universo. Quando  $N = \infty$  a fórmula (8) se reduz a

$$\sigma_x^2 = \sigma^2/n$$

utilizável no caso de extração com reposição (população infinita).

Se  $N$  é muito grande e  $n$  é pequeno em relação a  $N$  a fração  $(n-1)/(N-1)$  corresponde praticamente a  $n/N = p$ , de modo que a expressão anterior pode ser substituída por

$$\sigma_x^2 = (\sigma^2/n) (1 - p)$$

onde  $p$  representa a fração de amostragem da população. Seu inverso,  $N/n$ , fator pelo qual o valor total da amostra é multiplicado, corresponde ao fator de expansão. Observe-se, finalmente, que quando não há reposição a variância da média é menor que quando há reposição, posto que  $(N - n)/(N - 1)$  é sempre inferior à unidade.

#### VARIÂNCIA DAS MÉDIAS

Em vez de se considerar a população como um todo indecomponível, resultados mais satisfatórios podem ser obtidos subdividindo-o em grupos ou estratos. A Amostragem feita com base nesses estratos constitui a amostragem estratificada. Cumpre notar que esses estratos em que foi dividida a população, devem ser formados, tanto quanto possível, de elementos homogêneos, de modo a reduzir ao mínimo a variância dentro de cada um deles (variância interna), já que com a estratificação fica completamente eliminada a variância entre os

estratos (variância externa).

Considere-se, pois, uma população de tamanho  $N$ , subdividida em  $k$  estratos, contendo cada amostra  $N_i$  elementos. Representando por  $\bar{X}_i$  a média de cada estrato e por  $\bar{X}$  a média da população, a expressão desta será:

$$\bar{X} = \frac{\sum N_i \bar{X}_i}{\sum N_i} = (1/N)(N_1 \bar{X}_1 + N_2 \bar{X}_2 + \dots + N_k \bar{X}_k)$$

A melhor estimativa para a média da população é o valor  $\bar{x}$  que resulta ao se substituir na expressão anterior cada  $\bar{X}_i$  pela média  $\bar{x}_i$  de cada estrato obtida na amostra, isto é,

$$\bar{X} = (1/N) (N_1 \bar{x}_1 + N_2 \bar{x}_2 + \dots + N_k \bar{x}_k)$$

ou

$$\bar{X} = (N_1/N) \bar{x}_1 + (N_2/N) \bar{x}_2 + \dots + (N_k/N) \bar{x}_k$$

Fazendo  $a_i = N_i/N$  e substituindo acima, resulta:

$$\bar{X} = a_1 \bar{x}_1 + a_2 \bar{x}_2 + \dots + a_k \bar{x}_k$$

Sendo a variância da variável aleatória  $Z$  dada por

$$\sigma_z^2 = \sum a_i^2 \sigma_{x_i}^2$$

no caso de serem independentes as variáveis  $X_i$ , e substituindo,

$$\sigma_w^2, \text{ por } \sigma_x^2, a_i^2 \text{ por } N_i^2/N^2 \text{ e } \sigma_{x_i}^2 \text{ por } \sigma_{x_i}^2$$

vem

$$\sigma_x^2 = \sum (N_i^2/N^2) \sigma_{x_i}^2$$

Sendo a variância das médias  $\bar{X}_i$ , no caso de extração sem reposição, igual a

$$\sigma_{\bar{X}_i}^2 = (\sigma_i^2/n_i) (N_i - n_i)/(N_i - 1)$$

e substituindo na expressão anterior, resulta:

$$\sigma_x^2 = \sum_{i=1}^k (N_i^2/N) (N_i - n_i)/(N_i - 1) \sigma_i^2/n_i$$

ou

$$\sigma_x^2 = \sum_{i=1}^k (N_i^2) (1/N^2) (N_i - n_i)/(N_i - 1) (\sigma_i^2/n_i)$$

ou, finalmente,

$$\sigma_x^2 = 1/N^2 \sum_{i=1}^k (N_i^2) (N_i - n_i)/(N_i - 1) (\sigma_i^2/n_i) \quad (10)$$

expressão que fornece a variância das médias no caso de amostragem estratificada, na qual  $N$  representa o tamanho do universo;  $N_i$  o tamanho do estrato  $i$ ;  $n_i$  o número de elementos do estrato  $i$  e  $\sigma_i^2$  a variância do estrato  $i$ .

*Variância das médias no caso de amostragem estratificada com distribuição proporcional*

Consiste este processo em retirar os elementos da amostra proporcionalmente ao tamanho dos estratos em que foi dividido o universo.

De modo geral, deve-se ter a seguinte relação:

$$n_i/n = N_i/N \quad \text{ou} \quad n_i = (n/N) N_i \quad (11)$$

Subtraindo ambos os membros de  $N_i$ , resulta:

$$N_i - n_i = N_i - (n/N) N_i = N_i(1 - n/N) \quad (12)$$

Dividindo ambos os membros de (12) por (11):

$$(N_i - n_i)/n_i = N_i(1 - n/N)/(n/N)N_i \quad (13)$$

Dividindo ambos os termos do 2º membro e simplificando obtêm-se:

$$(N_i - n_i)/n_i = [(N - n)/N^2] / [(n/N^2)] = (N - n)/n \quad (14)$$

Substituindo-se  $N_i - 1$  por  $N_i$  na fórmula (11), o que não afetará sensivelmente o resultado, sobretudo se  $N$  for grande, obtêm-se:

$$\sigma_X^2 = (1/N^2) \sum N_i (N_i - n_i) \sigma_i^2 / N_i$$

ou

$$\sigma_X^2 = (1/N^2) \sum [N_i (N_i - n_i) / n_i] \sigma_i^2$$

ou, ainda, substituindo  $(N_i - n_i)/n$  pelo seu valor dado em (14):

$$\sigma_X^2 = 1/N^2 \sum [N_i (N - n) / N] \sigma_i^2$$

Passando  $1/N$  para dentro do somatório e retirando dele a fração  $(N - n)/n$ , que é constante, resulta, após simplificações:

$$\sigma_X^2 = [(N - n) / (Nn)] [\sum N_i \sigma_i^2 / \sum N_i^2] \quad (15)$$

expressão que fornece a variância das médias, no caso de amostragem proporcional mais facilmente que com o emprego da fórmula (11), porém, ainda suscetível de nova simplificação. Com efeito, substituindo-se o  $N$  que figura no denominador da 1ª fração por  $N - 1$ , ter-se-á:

$$\sigma_X^2 = [(N - n) / (N - 1)] \sum N_i \sigma_i^2 / \sum N_i \quad (1/n) \quad (16)$$

Sendo a variância total, no caso de amostragem estratificada com distribuição proporcional, igual à variância interna pode-se escrever:

$$\sum N_i \sigma_i^2 / \sum N_i = \sigma_i^2 \quad (17)$$

ou, ainda,

$$\sigma_x^2 = (N - n)/(N - 1) (\sigma_1^2/n) \quad (18)$$

valendo notar que essa fórmula foi deduzida para o caso de população finita (extração sem reposição). Para população infinita (extração com reposição), a 1ª fração do 2º membro desaparece e a expressão da variância das médias se reduz a:

$$\sigma_x^2 = \sigma_1^2/n \quad (19)$$

Comparando-se as expressões (18) e (19) com as equivalentes no caso de amostragem aleatória simples, onde

$$\sigma^2 = \sigma_1^2 + \sigma_E^2$$

constata-se que mediante a estratificação elimina-se a variância externa.

*Variância das médias no caso de amostragem estratificada com distribuição ótima*

Na ausência completa de informações a respeito da variância dos estratos em que foi subdividida a população, emprega-se a amostragem proporcional visto que proporcionará resultados mais satisfatórios, que a amostragem aleatória simples. Conhecendo-se a variância dos estratos, resulta mais eficaz adotar-se a amostragem estratificada com distribuição ótima. De acordo com esse critério o número  $n_i$  de elementos do estrato  $i$  será dado pela expressão:

$$n_i = n (N_i \sigma_i / \sum (N_i \sigma_i)) = (1/\sum N_i \sigma_i) N_i \sigma_i n \quad (20)$$

onde  $n$  representa o tamanho da amostra total.

Evidentemente, se os desvios padrões dos estratos forem iguais entre si, recai-se na amostragem proporcional.

Sendo  $\sum N_i \sigma_i$  constante, pode-se escrever

$$\alpha = 1/\sum N_i \sigma_i \text{ donde } n_i = \alpha N_i \sigma_i n \quad (21)$$

Substituindo  $n_i$  na fórmula geral da variância das médias dada em (10) resulta:

$$\begin{aligned} \sigma_x^2 &= 1/N^2 \sum N_i^2 (N_i - \alpha N_i \sigma_i n / (N_i - 1)) (\sigma_i^2 / \alpha N_i \sigma_i n) \\ \text{Substituindo } N_i - 1 &\text{ por } N_i \text{ e simplificando, vem} \\ \sigma_x^2 &= 1/N^2 [ (\sum N_i \sigma_i / \alpha n) - \sum N_i \sigma_i^2 ] \end{aligned} \quad (22)$$

Mas

$$\sum N_i \sigma_i / \alpha n = (1/\alpha) (\sum N_i \sigma_i)$$

Todavia, de (21) extrai-se

$$1/\alpha = \sum N_i \sigma_i$$

Portanto,

$$\sum N_i \sigma_i / \alpha n = \sum N_i \sigma_i (\sum N_i \sigma_i) n = (\sum N_i \sigma_i)^2 / n$$

Substituindo em (22) o valor acima obtido, após algumas simplificações, resulta:

$$\sigma_x^2 = (1/n) (\sum N_i \sigma_i / N)^2 - (1/N) (\sum N_i \sigma_i^2 / N)$$

Somando e subtraindo  $[(N - n)/Nn] [\sum N_i \sigma_i^2 / N]$  ao 2º termo da expressão anterior e substituindo  $N$  do 1º termo por  $\sum N_i$ , vem

$$\sigma_x^2 = 1/n - (\sum N_i \sigma_i / \sum N_i)^2 + (N - n)/Nn [(\sum N_i \sigma_i^2 / N)] - \\ - (N - n)/Nn [(\sum N_i \sigma_i^2 / N) - 1/N (\sum N_i \sigma_i^2 / N)]$$

ou, alterando a ordem,

$$\sigma_x^2 = (N - n)/Nn [(\sum N_i \sigma_i^2 / N)] + 1/n (\sum N_i \sigma_i^2)^2 / \sum N_i - \\ - \sum N_i \sigma_i^2 / N [(N - n)/Nn] - 1/N (\sum N_i \sigma_i^2) / N \quad (23)$$

Os dois últimos termos do 2º membro podem ser assim escritos:

$$- (\sum N_i \sigma_i^2) / N [(N - n)/Nn] + 1/N = - (\sum N_i \sigma_i^2) / N [(N - n)Nn] + \\ + n/Nn$$

ou

$$- [(\sum N_i \sigma_i^2) / N] (N/Nn) = - (\sum N_i \sigma_i^2) / Nn$$

Substituindo em (23), resulta: sucessivamente

$$\sigma_x^2 = (N - n)/Nn [(\sum N_i \sigma_i^2 / N)] + 1/n (\sum N_i \sigma_i) / \sum N_i^2 - N_i \sigma_i^2 / Nn$$

$$\sigma_x^2 = (N - n)/Nn [(\sum N_i \sigma_i^2) / N] \{ [\sum N_i \sigma_i^2 / Nn] - [1/n (\sum N_i \sigma_i / \sum N_i)]^2 \}$$

$$\sigma_x^2 = (N - n)/Nn [(\sum N_i \sigma_i^2 / N) - 1/n [\sum N_i \sigma_i^2 / \sum N_i - (\sum N_i \sigma_i / \sum N_i)^2]]$$

ou

$$\sigma_x^2(ot) = \sigma_x^2(prop) - \sigma_\sigma^2/n \quad (24)$$

onde se verifica que a variância das médias da distribuição ótima é menor do que a variância das médias da distribuição proporcional, notando-se, finalmente, que a expressão

$$\sigma_\sigma^2 = \sum N_i \sigma_i^2 / N_i - (\sum N_i \sigma_i / \sum N_i)^2 \quad (25)$$

fornece a variância dos desvios padrões dos diversos estratos em que foi subdividida a população.

Substituindo o 1º termo do 2º membro pelo seu valor dado em (16), vem

$$\sigma_x^2(ot) = (N - n)/(N - 1) [\sum N_i \sigma_i^2 / \sum N_i] / n - \sigma_\sigma^2/n$$

ou tendo em vista a expressão dada em (17):

$$\sigma_x^2(ot) = [(N - n)/(N - 1)] \sigma_I^2/n - \sigma_\sigma^2/n \quad (26)$$

no caso de extração sem reposição. Em se tratando de universo infinito

desaparece a 1ª fração do 2º membro e a expressão anterior se reduz a

$$\sigma_{\bar{x}}^2(\text{ot}) = (\sigma_1^2 - \sigma_0^2)/n \quad (27)$$

#### DIMENSIONAMENTO DO TAMANHO DA AMOSTRA

##### *Amostragem aleatória simples*

Denomina-se tamanho da amostra ao número de elementos que constituem a amostra. Demonstra-se que a média aritmética das médias de todas as amostras possíveis extraídas de uma população coincide com a média aritmética da própria população, resultado que se aplica a amostras de qualquer tamanho, extraídas com reposição ou sem reposição do elemento extraído, independentemente da forma da população.

No tocante à distribuição das médias das amostras é oportuno ressaltar dois aspectos: o de que sendo normal a população, a distribuição das médias também será normal, e o de que embora não sendo normal o universo ainda assim a distribuição das médias das amostras é aproximadamente normal, e tanto mais próxima da normalidade quanto maior for o tamanho da amostra.

Suponha-se que foram extraídas diversas amostras, suficientemente grandes, de um universo infinito. Sendo, nesse caso, normal a distribuição das médias pode-se utilizar na determinação do tamanho da amostra a própria curva normal de probabilidade. Tomando por base a aproximação fornecida por essa curva, facilmente se calculam as frações do total de amostras possíveis em que a média da amostra defira da média do universo de um, dois, três, etc. desvios padrões para mais ou para menos. Assim, por exemplo, sabe-se que cerca de 68,3% das amostras dão lugar a médias que diferem da média da população, para mais ou para menos, de um desvio padrão, cerca de 95,4% dão lugar a médias que diferem de 2 desvios padrões e 99,7% aproximadamente, dão lugar a média que diferem de 3 desvios padrões.

##### *Extração com reposição*

Sendo  $\mu$  a média da população, os resultados acima exemplificados podem ser simbolicamente representados assim:

$$\mu \pm \sigma ; \mu \pm 2\sigma ; \mu \pm 3\sigma$$

Desejando-se extrair uma amostra de uma população qualquer para estudar, por exemplo, a característica idade, de tal modo que 95% das amostras extraídas nas mesmas condições não difiram da média da população de mais ou menos 0,5 ano, escrever-se-á

$$1,96\sigma_{\bar{x}} \leq 0,5$$

De modo geral ter-se-á:

$$k\sigma_{\bar{x}} \leq E$$

em que E representa um valor qualquer predeterminado. Elevando-se ambos os membros ao quadrado tem-se

$$k^2 \frac{\sigma^2}{X} \leq E^2 \quad (28)$$

mas, sendo

$$\frac{\sigma^2}{X} = \sigma^2/n$$

vem

$$k^2 \sigma^2/n \leq E^2$$

donde

$$n \geq k^2 \sigma^2/E^2 \quad (29)$$

*Extração sem reposição*

Nesse caso, a variância das médias é dada por

$$\sigma^2_X = (\sigma^2/n) (N - n)/(N - 1)$$

a qual, substituída em (28), fornece:

$$k^2 \sigma^2/n [(N - n)/(N - 1)] \leq E^2$$

Multiplicando ambos os membros por n (N - 1), resulta, su  
cessivamente,

$$k^2 \sigma^2 (N - n) \leq E^2 n (N - 1)$$

$$k^2 \sigma^2 N - nk^2 \sigma^2 \leq E^2 n (N - 1)$$

$$E^2 n (N - 1) + nk^2 \sigma^2 \leq k^2 \sigma^2 N$$

donde, finalmente,

$$n \geq k^2 \sigma^2 N/E^2 (N - 1) + k^2 \sigma^2 \quad (30)$$

Conforme antes assinalado, a variância das médias no caso de extração sem reposição é menor do que a variância das médias no caso de extração com reposição, pois a fração (N - n)/(N - 1) é menor do que a unidade. Conseqüentemente, mediante uma amostragem aleatória simples, sem reposição, obtém-se a mesma precisão fornecida pela amostragem com reposição adotando-se uma amostra de tamanho menor, ou maior precisão para o mesmo tamanho da amostra.

*Amostragem estratificada com distribuição proporcional*

*Extração com reposição*

Para determinar o tamanho de uma amostra deve-se ter, se  
gundo a (28):

$$k^2 \frac{\sigma^2}{X} \leq E^2$$

Substituindo a variância da média pelo seu valor dado em (19), vem

$$k^2 \sigma^2_I/n \leq E^2$$

donde

$$n \geq k^2 \sigma_1^2 / E^2 \quad (31)$$

*Extração sem reposição*

Substituindo a variância da média pelo respectivo valor dado em (18), resulta, sucessivamente:

$$k^2 (N - n) / (N - 1) (\sigma_1^2 / n) \leq E^2$$

$$k^2 (N - n) \sigma_1^2 \leq E^2 (N - 1) n$$

$$k^2 \sigma_1^2 N - k^2 \sigma_1^2 n \leq E^2 (N - 1) n$$

$$k^2 \sigma_1^2 \leq n [E^2 (N - 1) + k^2 \sigma_1^2]$$

ou, finalmente,

$$n \geq (k^2 \sigma_1^2 N) / (E^2 (N - 1) + k^2 \sigma_1^2) \quad (32)$$

a qual, para N infinito, se transforma na expressão (31).

*Amostragem estratificada com distribuição ótima*

*Extração com reposição*

Substituindo a variância das médias pelo seu valor dado em (27), vem:

$$k^2 (\sigma_1^2 / n) - (\sigma_0^2 / n) \leq E^2$$

$$(k^2 \sigma_1^2 / n) - (k^2 \sigma_0^2 / n) \leq E^2$$

$$k^2 \sigma_1^2 - k^2 \sigma_0^2 \leq n E^2$$

ou

$$n \geq k^2 (\sigma_1^2 - \sigma_0^2) / E^2 \quad (33)$$

*Extração sem reposição*

Substituindo a variância das médias pelo seu valor dado em (26), vem

$$k^2 [(N - n) / (N - 1) (\sigma_1^2 / n) - \sigma_0^2 / n] \leq E^2$$

Efetuada a operação e multiplicando ambos os membros por  $n(N - 1)$ , resulta:

$$k^2 (N - n) \sigma_1^2 - k^2 \sigma_0^2 (N - 1) \leq E^2 n (N - 1)$$

donde, após algumas simplificações,

$$n \geq (k^2 \sigma_1^2 N) - (k^2 \sigma_0^2 (N - 1)) / E^2 (N - 1) + k^2 \sigma_1^2 \quad (34)$$

que para N infinito se transforma na expressão (33).

#### CONCLUSÕES SOBRE OS TIPOS DE AMOSTRAGEM

Em regra, na amostragem aleatória simples o número de elementos a extrair do universo para constituir a amostra é sempre

superior ao número correspondente na amostragem proporcional. Na amostragem estratificada com distribuição ótima o tamanho da amostra ainda é menor do que na proporcional. A redução considerável que as vezes se verifica ao passar de um tipo de amostragem para outro encontra explicação no fato de as médias dos estratos apresentarem uma grande variabilidade, bem como os desvios padrões. Conseqüentemente, a variância externa (entre os estratos) é muito grande. Ora, a variância total no caso da amostragem aleatória simples é igual à variância interna (dentro dos estratos) acrescida da variância externa, daí o tamanho da amostra ser relativamente elevado em comparação aos outros tipos de amostragem. Na estratificada com distribuição proporcional a variância total fica reduzida somente à variância interna, pois a estratificação elimina completamente a variância externa. O tamanho da amostra, graças à eliminação de uma das causas de variabilidade, sofre por isso uma redução sensível em relação à amostragem aleatória simples. Na amostragem estratificada com distribuição ótima, além de ficar eliminada a variância externa, ocorre também a eliminação da variância dos desvios padrões, o que conduz a um tamanho de amostra ainda menor do que nos demais tipos.

#### REFERÊNCIAS BIBLIOGRÁFICAS

1. COCHRAN, W.G. *Técnicas de Amostragem*. Fundo de Cultura S.A., Rio de Janeiro, 1965.
2. CRUZ, J. *Amostragem Estatística - Noções Básicas*. Empresa Gráfica Universitária S.A. Sergipe, 1978.
3. HOEL, P.G. *Introduction to Mathematical Statistics*. John Wiley, New York, 1962.
4. DIXON, W.J. & MASSEY, F.J.Jr. *Introduction to Statistical Analysis*. McGraw-Hill, New York, 1969.
5. GMURMAN, V.E. *Problemas em probabilidades e Estatística*. Editora Mir. Tradução, 1984.
6. TUKEY, I.W. *Exploratory data Analysis*. Reading, Addison Wesley, Massachusetts, 1977.

Recebido em novembro, 1987; aceito em dezembro, 1987.

