UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
CADERNOS DE MATEMÁTICA E ESTATÍSTICA
SÉRIE F: TRABALHO DE DIVULGAÇÃO

OS ESTIMADORES DA ESPERANÇA E DA VARIÂNCIA NA AMOSTRAGEM ALEATÓRIA SIMPLES COM E SEM REPOSIÇÃO

MARCO ANTÔNIO GIACOMELLI

SÉRIE F, Nº 13
PORTO ALEGRE, MARÇO 1997

OS ESTIMADORES DA ESPERANÇA E DA VARIÂNCIA NA AMOSTRAGEM ALEATÓRIA SIMPLES COM E SEM REPOSIÇÃO

MARCO ANTÔNIO GIACOMELLI

Endereço para correspondência:

Rua Alonso da Fonseca Coelho, 145

Bairro Agronomia

Porto Alegre - RS

CEP 91540 - 170

Telefone: (051) 981-73-59

RESUMO

O presente artigo tem por objetivo fazer uma discussão sobre os estimadores da esperança e da variância, respectivamente denotados por $\bar{X}$ e S^2, na amostragem aleatória simples. As propriedades, tais como esperança e variância, dos estimadores $\bar{X}$ e S^2 são apresentadas no contexto da amostragem com e sem reposição, sendo que a contribuição deste artigo encontra-se na obtenção, de uma maneira alternativa, da variância do estimador S^2 na amostragem sem reposição.

Unitermos: estimadores da esperança e da variância, amostragem aleatória simples.

0 - INTRODUÇÃO

Aqui neste trabalho nos concentraremos na amostragem aleatória simples, dando-se ênfase aos estimadores da esperança e da variância populacionais. O artigo divide-se em duas partes fundamentais: a primeira tratando da amostragem com reposição e a segunda da amostragem sem reposição. Na amostragem sem reposição faremos a dedução para a fórmula da variância do estimador não tendencioso $S^2 = \frac{N-1}{N}\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ da variância populacional σ^2, onde $(X_1, \cdots, X_n)$ é uma amostra aleatória de tamanho n de uma população de dimensão N e $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ é o estimador não tendencioso da esperança populacional μ. O método aqui empregado para deduzir a variância de S^2 é mais "natural" quando comparado, por exemplo, com o método de Sukhatme [5], isto é, neste artigo utilizaremos apenas os conceitos de momentos amostrais e populacionais, que são bastante conhecidos.

Cabe ainda salientar que o fato deste artigo tratar especificamente da amostragem aleatória simples não implica que esta deva ser preterivelmente escolhida para realizar um levantamento estatístico. Sabe-se que a amostragem aleatória simples apresenta algumas desvantagens com relação a outras técnicas, como por exemplo, a amostragem estratificada e a por conglomerados. A principal desvantagem da amostragem aleatória simples é quando a variabilidade da população, em relação a uma variável X, for grande. Então, neste caso, recomenda-se um processo de amostragem estratificada (ver Cochran [1]).

1 - O PROCESSO DE AMOSTRAGEM COM REPOSIÇÃO

Seja uma amostra aleatória $(X_1, \cdots, X_n)$ de uma população com esperança $E(X_1) \equiv \mu$ e variância $Var(X_1) \equiv \sigma^2 > 0$. A expressão "amostra aleatória" por convenção significa que as variáveis aleatórias $(X_1, \cdots, X_n)$ são identicamente distribuídas e independentes (no caso da amostragem sem reposição). Note que a denominação "amostra aleatória" é utilizada em Inferência Estatística exatamente com este significado. Aqui, nesta seção, diremos apenas "amostra aleatória" para nos referirmos ao processo de amostragem aleatória simples com reposição.

As demostrações das fórmulas que a seguir são apresentadas podem ser obtidas nas referências [1] e [3], por exemplo.

O estimador não tendencioso da esperança μ é dado por $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, e sabe-se que

$$E(\bar{X}) = \mu, \tag{1.1}$$

além de

$$Var(\bar{X}) = \frac{\sigma^2}{n}. \tag{1.2}$$

Para a variância populacional σ^2 é fato conhecido que $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ é o respectivo estimador não tendencioso. Aém disso,

$$E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = E\left(\sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2\right) = E\left(\sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2\right)$$

$$= \sum_{i=1}^{n} E(X_i - \mu)^2 - nE(\bar{X} - E(\bar{X}))^2 = n\left(\sigma^2 - Var(\bar{X})\right),$$

pois por (1.1), $E(\bar{X}) = \mu$. Como $(X_1, \cdots, X_n)$ são independentes, por (1.3) temos que

$$E(S^2) = \frac{1}{n-1}\left(n\sigma^2 - n\frac{\sigma^2}{n}\right) = \sigma^2.$$

Falta-nos ainda obter a fórmula da variância de S^2. Para tal, denotemos:

$$M_k' = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \quad \mu_k' = E(X_1^k), \quad M_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k \text{ e } \mu_k = E\left((X_1 - \mu)^k\right).$$

Em particular, $\mu_1' = E(X_1) = \mu$ e $\mu_2 = E\left((X_1-\mu)^2\right) = \sigma^2$. Notando que $S^2 = \left(\frac{n}{n-1}\right)\tilde{S}^2$, sendo $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, então faremos primeiro a obtenção de $Var(\tilde{S}^2)$. Lembrando que $Var(\tilde{S}^2) = E\left((\tilde{S}^2)^2\right) - \left(E(\tilde{S}^2)\right)^2$, comecemos pelo cálculo de $E(\tilde{S}^2)$. Uma vez que

$$\tilde{S}^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}^2 = M_2' - (M_1')^2,$$

então

$$E(\tilde{S}^2) = E\left(\frac{\sum_{i=1}^{n} X_i^2}{n}\right) - E(\bar{X}^2) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E(\bar{X}^2) = \mu_2' - E(\bar{X}^2) = \sigma^2 + \mu^2 - E(\bar{X}^2).$$

Mas $E(\bar{X}^2) = Var(\bar{X}) + \left(E(\bar{X})\right)^2 = \frac{\sigma^2}{n} + \mu^2$. Logo,

$$E(\tilde{S}^2) = \sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n} + \mu^2\right) = \sigma^2\left(\frac{n-1}{n}\right).$$

Agora, procedendo no cálculo de $E\left((\tilde{S}^2)^2\right)$, observemos que

$$E\left((\tilde{S}^2)^2\right) = E\left((M_2' - \bar{X}^2)^2\right) = E\left((M_2')^2\right) - 2E\left(M_2'\bar{X}^2\right) + E\left(\bar{X}^4\right).$$

Obtendo $E\left((M_2')^2\right)$:

$$E\left((M_2')^2\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} E(X_i^4) + 2\sum_{i<j}^{n} E(X_i^2X_j^2)\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} E(X_i^4) + 2\sum_{i<j}^{n} E(X_i^2)E(X_j^2)\right)$$

$$= \frac{1}{n^2}\left(n\mu_4' + 2\binom{n}{2}\mu_2'\mu_2'\right) = \frac{\mu_4' + (n-1)(\mu_2')^2}{n},$$

onde usamos o fato de $E(X_i^2 X_j^2) = E(X_i^2)E(X_j^2)$ pela independência de $(X_1, \cdots, X_n)$.

Obtendo $E\left(M_2' \bar{X}^2\right)$:

$$
\begin{aligned}
E\left(M_2'\bar{X}^2\right) &= \frac{1}{n^3}E\left((X_1^2+\cdots+X_n^2)(X_1+\cdots+X_n)^2\right)\\
&= \frac{1}{n^3}E\left(\sum_{i=1}^{n}X_i^4 + 2\sum_{i<j}^{n}X_i^2X_j^2 + 2\sum_{i<j}^{n}X_i^3X_j + 2\sum_{i<j}^{n}X_iX_j^3 + 2\sum_{i<j<k}^{n}X_i^2X_jX_k\right.\\
&\left.+2\sum_{i<j<k}^{n}X_iX_j^2X_k + 2\sum_{i<j<k}^{n}X_iX_jX_k^2\right)\\
&= \frac{1}{n^3}\left(n\mu_4' + \binom{n}{2}(\mu_2')^2 + 4\binom{n}{2}\mu_3'\mu + 6\binom{n}{3}\mu_2'\mu^2\right)\\
&= \frac{1}{n^2}\left(\mu_4' + (n-1)(\mu_2')^2 + 2(n-1)\mu_3'\mu + (n-1)(n-2)\mu_2'\mu^2\right).
\end{aligned}
$$

Obtendo $E\left(\bar{X}^4\right)$:

$$
\begin{aligned}
E\left(\bar{X}^4\right) &= \frac{1}{n^4}E\left(\left(\sum_{i=1}^{n}(X_i-\mu)+n\mu\right)^4\right)\\
&= \frac{1}{n^4}\left(E\left(\left(\sum_{i=1}^{n}(X_i-\mu)\right)^4\right) + (n\mu)^4 + 6(n\mu)^2E\left(\left(\sum_{i=1}^{n}(X_i-\mu)\right)^2\right)\right.\\
&\left.+4n\mu E\left(\left(\sum_{i=1}^{n}(X_i-\mu)\right)^3\right) + (n\mu)^3E\left(\sum_{i=1}^{n}(X_i-\mu)\right)\right).
\end{aligned}
$$

Como, ver Murteira [3] - pág. 58,

$$E\left(\left(\sum_{i=1}^{n}(X_i-\mu)\right)^4\right) = n\mu_4 + 3n(n-1)(\mu_2)^2;$$

$$E\left(\left(\sum_{i=1}^{n}(X_i-\mu)\right)^3\right) = n\mu_3;$$

$$E\left(\left(\sum_{i=1}^{n}(X_i-\mu)\right)^2\right) = n\mu_2;$$

$$E\left(\left(\sum_{i=1}^{n}(X_i-\mu)\right)\right) = 0,$$

então $E(\bar{X}^4) = \mu^4 + \frac{\mu_4+3(n-1)\sigma^4}{n^3} + \frac{6\mu^2\sigma^2}{n} + \frac{4\mu_3\mu}{n^2}$.

Sem perda de generalidade, podemos supor que $\mu = 0$. Assim,

$$E\left((M_2')^2\right) = \frac{\mu_4' + (n-1)(\mu_2')^2}{n};$$

$$E\left(M_2'\bar{X}^2\right) = \frac{\mu_4' + (n-1)(\mu_2')^2}{n^2};$$

$$E\left(\bar{X}^4\right) = \frac{\mu_4' + 3(n-1)(\mu_2')^2}{n^3};$$

$$E\left(\tilde{S}^2\right) = \mu_2'\left(\frac{n-1}{n}\right).$$

Logo,

$$Var\left(\tilde{S}^2\right) = \frac{\mu_4' + (n-1)(\mu_2')^2}{n} - 2\left(\frac{\mu_4' + (n-1)(\mu_2')^2}{n^2}\right) + \frac{\mu_4' + 3(n-1)(\mu_2')^2}{n^3} - (\mu_2')^2\left(\frac{n-1}{n}\right)^2$$

$$= \frac{\mu_4'(n-1)^2 - (n-1)(n-3)(\mu_2')^2}{n^3},$$

e portanto,

$$Var(S^2) = \frac{n^2}{(n-1)^2} Var(\tilde{S}^2) = \frac{1}{n}\left(\mu_4' - \frac{n-3}{n-1}(\mu_2')^2\right).$$

No caso em que $\mu \neq 0$,

$$Var(S^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right). \tag{1.4}$$

2 - O PROCESSO DE AMOSTRAGEM SEM REPOSIÇÃO

No processo de amostragem sem reposição não dispomos da propriedade de independência das variáveis aleatórias, e por isso, os cálculos de valores esperados são mais complicados.

Para a amostragem sem reposição, o estimador não tendencioso de $E(X_1) = \mu$ é $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Analogamente ao caso com reposição, prova-se que $E(\bar{X}) = \mu$. Por outro lado, para calcularmos $Var(\bar{X})$ iremos supor que N, denotando o tamanho da população, seja finito. Como na amostragem sem reposição $Cov(X_i, X_j) \neq 0$, para $\forall i < j$, então teremos de calcular a covariância. Mas como

$$Cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j) = E(X_iX_j) - \mu^2 = E(X_iX_j) - \left(E(X_1^2) - \sigma^2\right),$$

então agora procederemos na obtenção de $E(X_iX_j)$. Sendo $\{x_1, \cdots, x_N\}$ as unidades populacionais, e relembrando que $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ e $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$, então

$$\begin{aligned}
E(X_iX_j) &= \frac{1}{\binom{N}{2}}(x_1x_2 + x_1x_3 + \cdots + x_{N-1}x_N) = \frac{1}{N(N-1)}(2x_1x_2 + 2x_1x_3 + \cdots + 2x_{N-1}x_N) \\
&= \frac{1}{N(N-1)}(x_1x_2 + x_2x_1 + \cdots + x_{N-1}x_N + x_Nx_{N-1}) \\
&= \frac{1}{N(N-1)}\left((x_1 + \cdots + x_N)^2 - (x_1^2 + \cdots + x_N^2)\right) \\
&= \frac{1}{N(N-1)}\left(N^2\mu^2 - \sum_{i=1}^{N} x_i^2\right) = \frac{N^2\mu^2}{N(N-1)} - \frac{1}{N-1}E(X_1^2).
\end{aligned}$$

Daí, $Cov(X_i, X_j) = -\frac{N}{N(N-1)}\sigma^2$. Uma vez que as variáveis aleatórias são identicamente distribuídas, $Cov(X_i, X_j)$ é a mesma para qualquer $i < j$. Assim, por (1.2),

$$\text{(2.1)} \qquad Var(\bar{X}) = \frac{1}{n^2}\left(n\sigma^2 + 2\binom{n}{2}\left(-\frac{N}{N(N-1)}\sigma^2\right)\right) = \frac{N-n}{N-1}\frac{\sigma^2}{n}.$$

Note que em (2.1), $\frac{N-n}{N-1}$ é denominada de fator de correção para populações finitas. Quando $N \to \infty$, $\frac{N-n}{N-1}$ tende a 1. Isto implica que não haverá diferença significativa nas variâncias dos processos de amostragem com e sem reposição, para N grande.

O estimador não tendencioso da variância populacional é $S^2 = \frac{N-1}{N}\frac{n}{n-1}\tilde{S}^2$, sendo $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. De fato, como

$$E(\tilde{S}^2) = E(X_1^2) - E(\bar{X}^2) = \sigma^2 + \mu^2 - \left(Var(\bar{X}) + (E(\bar{X}))^2\right)$$

$$(2.2) \qquad = \sigma^2 + \mu^2 - \left(\frac{N-n}{N-1}\frac{\sigma^2}{n} + \mu^2\right) = \frac{N}{N-1}\frac{n-1}{n}\sigma^2,$$

então $E(S^2) = \sigma^2$.

No cálculo de $Var(\tilde{S}^2)$ utilizaremos o mesmo raciocínio do processo com reposição, isto é, iremos aproveitar a relação

$$(2.3) \qquad Var(\tilde{S}^2) = E\left((\tilde{S}^2)^2\right) - \left(E(\tilde{S}^2)\right)^2.$$

Sabe-se que

$$(2.4) \qquad E\left((\tilde{S}^2)^2\right) = E\left((M_2')^2\right) - 2E\left(M_2'\bar{X}^2\right) + E\left(\bar{X}^4\right).$$

Obtendo $E\left((M_2')^2\right)$:

$$E\left((M_2')^2\right) = \frac{1}{n^2}E\left(\left(\sum_{i=1}^{n} X_i^2\right)^2\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} E(X_i^4) + 2\sum_{i<j}^{n} E(X_i^2X_j^2)\right).$$

Mas uma vez que $E(X_i^2X_j^2)$ é a mesma para $i \neq j = 1, \cdots, n$ (pelo fato das v'as. serem identicamente distribuídas) então basta obter $E(X_1^2X_2^2)$. Sendo assim

$$E(X_i^2X_j^2) = \frac{1}{\binom{N}{2}}\left(x_1^2x_2^2 + x_1^2x_3^2 + \cdots + x_{N-1}^2x_N^2\right) = \frac{1}{N(N-1)}\left(2x_1^2x_2^2 + 2x_1^2x_3^2 + \cdots + 2x_{N-1}^2x_N^2\right)$$

$$= \frac{1}{N(N-1)}\left(\left(x_1^2 + x_2^2 + \cdots + x_N^2\right)^2 - \left(x_1^4 + x_2^4 + \cdots + x_N^4\right)\right)$$

$$= \frac{1}{N(N-1)}\left(N^2\left(\frac{\sum_{i=1}^{N} x_i^2}{N}\right)^2 - \sum_{i=1}^{N} x_i^4\right) = \frac{N\left(E(X_1^2)\right)^2}{N-1} - \frac{E(X_1^4)}{N-1}$$

$$(2.5) \qquad = \frac{N(\mu_2')^2 - \mu_4'}{N-1}.$$

Portanto,

$$E\left((M_2')^2\right) = \frac{1}{n^2}\left(n\mu_4' + 2\binom{n}{2}\frac{N(\mu_2')^2 - \mu_4'}{N-1}\right) = \frac{\mu_4'(N-n) + (\mu_2')^2(n-1)N}{n(N-1)}.$$

Obtendo $E\left(M_2'\bar{X}^2\right)$:

$$E\left(M_2'\bar{X}^2\right) = \frac{1}{n^3}E\left(\left(\sum_{i=1}^{n} X_i^2\right)\left(\sum_{i=1}^{n} X_i\right)^2\right) = \frac{1}{n^3}E\left(\sum_{i=1}^{n} X_i^4 + 2\sum_{i<j}^{n} X_i^2X_j^2 + 2\sum_{i<j}^{n} X_i^3X_j + 2\sum_{i<j}^{n} X_iX_j^3\right.$$

$$\left. +2\sum_{i<j<k}^{n} X_i^2X_jX_k + 2\sum_{i<j<k}^{n} X_iX_j^2X_k + 2\sum_{i<j<k}^{n} X_iX_jX_k^2\right)$$

$$= \frac{1}{n^3}\left(nE(X_1^4) + 2\binom{n}{2}E(X_1^2X_2^2) + 2\binom{n}{2}\left(E(X_1^3X_2) + E(X_1X_2^3)\right)\right.$$

$$\left. +2\binom{n}{3}\left(E(X_1^2X_2X_3) + E(X_1X_2^2X_3) + E(X_1X_2X_3^2)\right)\right).$$

Uma vez que

$$E(X_1^3X_2) + E(X_1X_2^3) = \frac{1}{\binom{N}{2}}\left(x_1^3x_2 + x_1^3x_3 + \cdots + x_{N-1}^3x_N\right) + \frac{1}{\binom{N}{2}}\left(x_1x_2^3 + x_1x_3^3 + \cdots + x_{N-1}x_N^3\right)$$

$$= \frac{(x_1^3 + x_2^3 + \cdots + x_N^3)(x_1 + x_2 + \cdots + x_N) - (x_1^4 + x_2^4 + \cdots + x_N^4)}{\binom{N}{2}}$$

$$= \frac{2x_1^3}{N-1}\frac{\sum_{i=1}^N x_i}{N} + \frac{2x_2^3}{N-1}\frac{\sum_{i=1}^N x_i}{N} + \cdots + \frac{2x_N^3}{N-1}\frac{\sum_{i=1}^N x_i}{N} - \frac{2}{N-1}\frac{\sum_{i=1}^N x_i^4}{N}$$

$$= \frac{2N}{N-1}\frac{\sum_{i=1}^N x_i}{N}\frac{\sum_{i=1}^N x_i^3}{N} - \frac{2}{N-1}\frac{\sum_{i=1}^N x_i^4}{N}$$

$$= \frac{2N}{N-1}E(X_1)E(X_1^3) - \frac{2}{N-1}E(X_1^4) = \frac{2N\mu\mu_3' - 2\mu_4'}{N-1} \tag{2.6}$$

e

$$
\begin{aligned}
E(X_1^2X_2X_3)+E(X_1X_2^2X_3)+E(X_1X_2X_3^2) &= \frac{1}{\binom{N}{3}}\left(x_1^2x_2x_3+x_1^2x_2x_4+\cdots+x_{N-2}^2x_{N_1}x_N\right)\\
&+\frac{1}{\binom{N}{3}}\left(x_1x_2^2x_3+x_1x_2^2x_4+\cdots+x_{N-2}x_{N_1}^2x_N\right)\\
&+\frac{1}{\binom{N}{3}}\left(x_1x_2x_3^2+x_1x_2x_4^2+\cdots+x_{N-2}x_{N_1}x_N^2\right)\\
&=\frac{6}{N(N-1)(N-2)}\left(\left(x_1^2+\cdots+x_N^2\right)\left(x_1x_2+\cdots+x_{N-1}x_N\right)\right.\\
&\left.-\left(x_1^3+\cdots+x_N^3\right)\left(x_1+\cdots+x_N\right)+\left(x_1^4+\cdots+x_N^4\right)\right)\\
&=\frac{3\left(\left(\sum_{i=1}^N x_i\right)^2-\sum_{i=1}^N x_i^2\right)\left(\sum_{i=1}^N x_i^2\right)}{N(N-1)(N-2)}\\
&-\frac{6\left(\sum_{i=1}^N x_i^3\right)\left(\sum_{i=1}^N x_i\right)}{N(N-1)(N-2)}+\frac{6\sum_{i=1}^N x_i^4}{N(N-1)(N-2)}\\
&=\frac{3\left(N^2\mu^2\mu_2'-N(\mu_2')^2\right)-6N\mu_3'\mu+6\mu_4'}{(N-1)(N-2)}, \qquad (2.7)
\end{aligned}
$$

então

$$
\begin{aligned}
E\left(M_2'\bar{X}^2\right) &= \frac{\mu_4'}{n^2}+\frac{(n-1)\left(N(\mu_2')^2-\mu_4'\right)+2(n-1)\left(N\mu_3'\mu-\mu_4'\right)}{n^2(N-1)}\\
&+\frac{(n-1)(n-2)\left(N^2\mu^2\mu_2'-N(\mu_2')^2-2N\mu_3'\mu+2\mu_4'\right)}{n^2(N-1)(N-2)}.
\end{aligned}
$$

Obtendo $E\left(\bar{X}^4\right)$:

$$E\left(\bar{X}^4\right)=\frac{1}{n^4}E\left(\left(\sum_{i=1}^{n}X_i\right)^4\right)=\frac{1}{n^4}E\left(\sum_{i=1}^{n}X_i^4+6\sum_{i<j}^{n}X_i^2X_j^2+4\sum_{i<j}^{n}X_i^3X_j+4\sum_{i<j}^{n}X_iX_j^3\right.$$

$$\left.+12\sum_{i<j<k}^{n}X_i^2X_jX_k+12\sum_{i<j<k}^{n}X_iX_j^2X_k+12\sum_{i<j<k}^{n}X_iX_jX_k^2+24\sum_{i<j<k<l}^{n}X_iX_jX_kX_l\right)$$

$$=\frac{1}{n^4}\left(nE(X_1^4)+6\binom{n}{2}E(X_1^2X_2^2)+4\binom{n}{2}\left(E(X_1^3X_2)+E(X_1X_2^3)\right)\right.$$

$$\left.+12\binom{n}{3}\left(E(X_1^2X_2X_3)+E(X_1X_2^2X_3)+E(X_1X_2X_3^2)\right)+24\binom{n}{4}E(X_1X_2X_3X_4)\right).$$

Observando que por (2.5), (2.6) e (2.7) já tem-se $E(X_1^2X_2^2)$, $E(X_1^3X_2)+E(X_1X_2^3)$ e $E(X_1^2X_2X_3)$ $+E(X_1X_2^2X_3)$ $+E(X_1X_2X_3^2)$, respectivamente, resta ainda calcular $E(X_1X_2X_3X_4)$. Sendo assim,

$$E\left(X_1X_2X_3X_4\right)=\frac{x_1x_2x_3x_4+\cdots+x_{N-3}x_{N-2}x_{N-1}x_N}{\binom{N}{4}}$$

$$=\frac{4}{N(N-1)(N-2)(N-3)}\left(\frac{1}{4}\left(\sum_{i=1}^{N}x_i\right)^4-\frac{3}{2}\left(\sum_{i=1}^{N}x_i\right)^2\left(\sum_{i=1}^{N}x_i^2\right)+\frac{3}{4}\left(\sum_{i=1}^{N}x_i^2\right)^2\right.$$

$$\left.+2\left(\sum_{i=1}^{N}x_i^3\right)\left(\sum_{i=1}^{N}x_i\right)--\frac{3}{2}\sum_{i=1}^{N}x_i^4\right)\Bigg)$$

$$=\frac{4}{N(N-1)(N-2)(N-3)}\left(\frac{N^4\mu^4}{4}-\frac{3N^3\mu^2\mu_2'}{2}+\frac{3N^2(\mu_2')^2}{4}+2N^2\mu_3'\mu-\frac{3N\mu_4'}{2}\right)$$

$$=\frac{1}{(N-1)(N-2)(N-3)}\left(N^3\mu^4-6N^2\mu^2\mu_2'+3N(\mu_2')^2+8N\mu_3'\mu-6\mu_4'\right).$$

Logo,

$$E\left(\bar{X}^4\right) = \frac{\mu_4'}{n^3} + \frac{(n-1)\left(3N(\mu_2')^2 - 7\mu_4' + 4N\mu_3'\mu\right)}{n^3(N-1)}$$

$$+ \frac{6(n-1)(n-2)\left(N^2\mu^2\mu_2' - N(\mu_2')^2 - 2N\mu_3'\mu + 2\mu_4'\right)}{n^3(N-1)(N-2)}$$

$$+ \frac{(n-1)(n-2)(n-3)\left(3N(\mu_2')^2 + 8N\mu_3'\mu + N^3\mu^4 - 6N^2\mu_2'\mu^2 - 6\mu_4'\right)}{n^3(N-1)(N-2)(N-3)}.$$

Para obter $E\left((\tilde{S}^2)^2\right)$ podemos novamente supor, sem perda de generalidade, que $\mu = 0$. Assim,

$$E\left((M_2')^2\right) = \frac{\mu_4'(N-n) + (\mu_2')^2(n-1)N}{n(N-1)};$$

$$E\left(M_2'\bar{X}^2\right) = \frac{\mu_4'}{n^2} + \frac{(n-1)\left(N(\mu_2')^2 - 3\mu_4'\right)}{n^2(N-1)} + \frac{(n-1)(n-2)\left(2\mu_4' - N(\mu_2')^2\right)}{n^2(N-1)(N-2)};$$

$$E\left(\bar{X}^4\right) = \frac{\mu_4'}{n^3} + \frac{(n-1)\left(3N(\mu_2')^2 - 7\mu_4'\right)}{n^3(N-1)} + \frac{6(n-1)(n-2)\left(2\mu_4' - N(\mu_2')^2\right)}{n^3(N-1)(N-2)}$$

$$+ \frac{(n-1)(n-2)(n-3)\left(3N(\mu_2')^2 - 6\mu_4'\right)}{n^3(N-1)(N-2)(N-3)}.$$

Mas

$$Var(S^2) = \left(\frac{N-1}{N}\right)^2 \left(\frac{n}{n-1}\right)^2 Var(\tilde{S}^2),$$

e utilizando (2.2), (2.3) e (2.4) chega-se, após algumas simplificações, a

$$Var(S^2) = \frac{\mu_4'(N-1)\left((N-2)(N-3)(n-1) - (N-1)(n-2)(n-3)\right)}{n(n-1)N(N-2)(N-3)}$$

$$+ (\mu_2')^2(N-1)\left(\frac{(n^2-2n+3)(N-2)(N-3) + (n-2)(n-3)(2N-3)}{n(n-1)N(N-2)(N-3)} - \frac{1}{N-1}\right).$$

De uma maneira geral,

$$Var(S^2) = \frac{\mu_4(N-1)\left((N-2)(N-3)(n-1)-(N-1)(n-2)(n-3)\right)}{n(n-1)N(N-2)(N-3)}$$

$$(2.8) \qquad + \sigma^4(N-1)\left(\frac{(n^2-2n+3)(N-2)(N-3)+(n-2)(n-3)(2N-3)}{n(n-1)N(N-2)(N-3)} - \frac{1}{N-1}\right).$$

Em (2.8), $\lim_{N\to\infty} Var(S^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right)$, e comparando com (1.4) vemos que para populações infinitas (ou suficientemente grandes) a variância do estimador S^2 é equivalente àquela no processo com reposição. Dessas considerações concluímos que mesmo quando o processo sem reposição é utilizado na prática em populações muito grandes, pode-se aplicar as fórmulas do processo com reposição, que são mais simplificadas.

3 - REFERÊNCIAS

[1] **Cochran, W. G. (1965).** *Sampling Techniques.* John Wiley & Sons, 2nd edition, New York.

[2] **Kish, L. (1965).** *Survey Sampling.* John Wiley & Sons, New York.

[3] **Murteira, B. J. F. (1990).** *Probabilidades e Estatística* , Vol. 2., McGraw-Hill de Portugal, 2ª edição, Lisboa.

[4] **Sudman, S. (1976).** *Applied Sampling.* Academic Press, London.

[5] **Sukhatme, P. V. (1963).** *Teoría de Encuestas por Muestro con Aplicaciones,* Fondo de Cultura Económica, 2ª edición, México.

Publicações do Instituto de Matemática da UFRGS
Cadernos de Matemática e Estatística

Série F: Trabalho de Divulgação

1. Nubem A. C. Medeiros e Jaime B. Ripoll - Superfícies Invariantes - MAI/90
2. Sídia M. C. Jacques - Análise de Correspondência: Aplicações em Genética - DEZ/91
3. Jandyra M. G. Fachel et al - Correspondence Analysis Aplied to Ethnographic Data: Case Examples - JAN/92
4. João Riboldi e Dinara W. X. Fernandez - Análise de Observações Repetidas através de Contrastes no Tempo - JUL/92
5. M. Teresa Albanese e Martin Knott - Twomiss: a Computer Program for Fitting a One or Two-factor Logit-probit Latent Variable Model to Binary Data when Observations May be Missing - JUL/92
6. Vera C. G. Carneiro - Retrato Atual do Curso de Licenciatura em Matemática da UFRGS e Plano de Ações Renovadoras - JUL/92
7. Suzi A. Camey e Jandyra M. G. Fachel - A Atuação dos Deputados Federais Gaúchos: Similaridades e Dissimilaridades em Relação as Votações mais Importantes no Período 91-94 - DEZ/94
8. Jandyra M. G. Fachel - Identificação de Areas Prioritárias para Desenvolvimento de Políticas Relativas à Situação da Infância de 0 a 6 anos no Rio Grande do Sul - DEZ/94
9. Dinara W. X. Fernandez, José L. D. Ribeiro e Ivo G. L. Wagner - Análise de uma Linha de Produção Envolvendo Políticas de Manutenção em Função da Criticidade dos Equipamentos: Simulação de Monte Carlo no Software XCELL+ - JUN/95

10. Marco A. Giacomelli - Grandes Desvios em Testes de Hipóteses - AGO/95
11. Marco A. Giacomelli - Modelos Log-lineares Aplicados a Estudos Epidemiológicos - SET/95
12. Dinara W. X. Fernandez - Profissão Estatístico - OUT/95
13. Marco A. Giacomelli - Os Estimadores da Esperança e da Variância na Amostragem Aleatória Simples com e sem Reposição MAR/97