Aspectos da teoria da amostragem (*)

por Rui João Baptista Soares
Laboratório de Física e Engenharia Nucleares
Sacavém — Portugal

A amostragem tem por finalidade principal a avaliação para um todo — o universo — de características a partir de uma parte — a amostra. Como trabalho estatístico que é processa-se essencialmente nas seguintes fases:

a) recolha de dados;
b) ordenação;
c) apresentação de resultados;
d) análise e interpretação dos resultados.

Tendo especificado o problema adopta-se um modelo matemático e simultâneamente métodos de amostragem que permitam, dada a impossibilidade de utilizar toda a informação, seleccionar elementos no universo.

Torna se então necessário formular um método M que será *aleatório* quando, pressupondo a existência da função condicional

$$F_i(x \mid x_1, \cdots, x_i) = P(X'_{i+1} < x \mid X'_1 = \\ = x_1, \cdots, X'_i = x_i) \quad i = 1, 2, \cdots$$

se tiver

1. $$P(X_1 < x) = F(x)$$

para todo o x

2. $$P(X_{i+1} < x \mid X_1 = x_1, \cdots, X_i = x_i) = \\ = F_i(x \mid x_1, \cdots, x_i) \quad i = 1, \cdots, n-1$$

quaisquer que sejam $x_1, \cdots, x_i$.

As amostras obtidas pelo método M dizem-se *aleatórias*. Se as variáveis aleatórias $X_1, \cdots, X_n$ são independentes a amostra é *simples*.

Podemos qualificar um método de *aceitável* quando reunir os seguintes requisitos:

A) custo reduzido;
B) rapidez e alcance suficientes;
C) precisão aceitável.

Posteriormente procede-se à optimização do método escolhido minimizando as funções

$$F_1 = C_0(\rho) + \lambda \sigma^2(\rho)$$

$$F_2 = \sigma_0(\rho) + \lambda C(\rho)$$

consoante se pretende que:

1) a variância σ^2 seja mínima para um custo prèviamente fixado;
2) o custo C seja mínimo para uma precisão exigida.

De entre os métodos probabilísticos os mais utilizados são:

1. Método elementar

Consiste em:

E_1) fraccionar o espaço num número finito ou infinito de unidades de amostra por forma que todo e qualquer elemento do universo pertença a uma única unidade;

E_2) estabelecer uma correspondência biunívoca entre cada unidade e uma bola de uma urna de Bernoulli;

(*) Este artigo contém os aspectos mais importantes de um trabalho sobre teoria de amostragem apresentado no Seminário de Estatística e Automática (5.° Ano de Matemática Aplicada de Lisboa) em 1968-1969.

E_3) dar um natural n a que daremos a designação de *volume da amostra;*

E_4) efectuar n tiragens (com reposição) da urna de Bernoulli e anotar sempre o número da bola extraída;

E_5) a amostra compõe-se exactamente das n unidades correspondendo às n bolas extraídas.

Este método pode ser utilizado em circunstâncias mais vantajosas desde que se disponha de uma *tabela de números aleatórios;* também nalguns casos é substituído pelo método das tiragens exaustivas em que E_4) dá lugar a:

E_4') efectuar n tiragens (sem reposição) da urna de Bernoulli e anotar sempre o número da bola extraída.

Em qualquer dos casos tem-se

$$E[\bar{x}] = m$$

caso particular do

Teorema 1. *Seja* $x_1, \cdots, x_n$ *uma amostra aleatória de uma população com densidade* $f(x)$. *A esperança matemática do momento central de ordem* r *da amostra é igual ao momento central, da mesma ordem, da população. Tem-se portanto*

$$E[m_r'] = \mu_r'.$$

Nas condições referidas, desde que a variância da população seja finita tem-se

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}.$$

Para calcular a variância da média, no caso de tiragens exaustivas associe-se a cada bola i da urna, uma variável aleatória I_i definida por

$$I_i = \begin{cases} 0 & \text{se a bola } i \text{ ficou na urna} \\ 1 & \text{» » » } i \text{ foi extraída.} \end{cases}$$

com probabilidades

$$P(I_i = 0) = 1 - \frac{n}{N}$$

$$P(I_i = 1) = \frac{n}{N}$$

sendo N o efectivo da população.

Atendendo a

$$P[(I_i \cap I_j) = 1] = \frac{N(n-1)}{N(N-1)}$$

e

$$(1.1) \qquad \sum_{i,j} x_i x_j = \left(\sum_i x_i\right)^2 - \sum_i x_i^2$$

resulta finalmente

$$\sigma_{\bar{x}_a}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} = \sigma_{\bar{x}}^2 \cdot (1-f);$$

$$\text{com } f = \frac{n-1}{N-1}$$

Sendo a *fracção de amostra* f desprezável segue-se que

$$\sigma_{\bar{x}_a}^2 \simeq \sigma_{\bar{x}}^2$$

isto é, os métodos são pràticamente equivalentes.

2. Amostragem sistemática

Suponhamos que é possível numerar as unidades de uma população; o método consiste em:

S_1) dar um número natural r;

S_2) tomar aleatòriamente uma unidade v_1 compreendida nas r primeiras unidades da população;

S_3) tomar as unidades $u_i = v_1 + kr$, $k \in \mathbb{N}$

S_4) a amostra é constituída pelas unidades u_j.

Admitamos que o conjunto dos valores observados se encontram dispostos sob a forma de matriz

$$X=[x_{ij}] \qquad \begin{matrix} 1\leq i\leq m \\ 1\leq j\leq n \end{matrix} \qquad n\geqslant 2$$

e consideremos as médias $\bar{x}_{i.}$, $\bar{x}_{.j}$ e $\bar{x}_{..}$ dos valores de linha, coluna e do quadro; sejam ainda σ^2, σ_1^2 e σ_2^2 as variâncias dos elementos do quadro, e das médias de linha e de coluna.

Pode então estabelecer-se a seguinte igualdade

$$(2.1) \qquad \sigma_1^2=\frac{\sigma^2}{n}+\frac{2}{n^2}\sum_{j<j_1} c_{jj_1}$$

onde c_{jj_1} é a covariância dos elementos correspondentes nas colunas j e j_1. A média destas quantidades é:

$$(2.2) \qquad \bar{c}=\frac{2}{n(n-1)}\sum_{j<j_1} c_{jj_1}.$$

Chamaremos *correlação interna* à grandeza $\bar{\rho}$ definida por

$$(2.3) \qquad \bar{\rho}=\frac{\bar{c}}{\sigma^2}.$$

De (2.1), (2.2) e (2.3) resulta a fórmula fundamental

$$(2.4) \qquad \sigma_1^2=\frac{\sigma^2}{n}[1+(n-1)\bar{\rho}]$$

que permite concluir

$$(2.5) \qquad \frac{1}{1-n}\leq\bar{\rho}\leq 1.$$

Tome-se a variância

$$(2.6) \qquad \sigma'^2=\frac{\sigma^2}{n}\cdot\frac{mn-n}{mn-1}$$

de uma média obtida numa amostra aleatória simples de efectivo n. Da comparação de (2.4) e (2.6) sai

$$(2.7) \qquad \sigma_1^2\lesseqgtr\sigma'^2 \iff \bar{\rho}\lesseqgtr\frac{1}{1-mn}$$

que permite ver quando é que a amostragem sistemática é mais precisa, de igual precisão ou menos precisa que a amostragem feita pelo método elementar.

Poder-se-ia ver [3] que a igualdade corresponde a uma distribuição ao acaso das unidades entre as linhas.

3. Amostragem por grupos.

Consideremos uma população constituída por m grupos de n_i elementos cada, $(1\leq i\leq m)$ onde se pretende estudar a propriedade X tomando o valor x_{ij} no elemento número j do grupo i.

G_1) dar um número natural $(\mu<m)$;

G_2) tomar aleatòriamente μ grupos;

G_3) a amostra será constituída por

$$n=\sum_{i=1}^{\mu} n_i$$

elementos.

A tiragem dos grupos é, nalguns casos, feita com probabilidades proporcionais ao efectivo do grupo.

Admitamos que os grupos têm o mesmo efectivo e calculemos a média a partir da amostra

$$\overset{*}{x}_{..}=\frac{1}{\mu n}\sum_{i=1}^{\mu}\sum_{j=1}^{n} x_{ij}$$

que é uma estimação centrada de $\bar{x}_{..}$.

Por considerações análogas às feitas em 2. chega-se a

$$(3.1)\quad \sigma^2_{\bar{x}^*} = \frac{\sigma^2(m-\mu)}{\mu n(m-1)}\cdot[1+(n-1)\bar{\rho}].$$

Como, no caso presente

$$(3.2)\qquad \sigma'^2 = \frac{\sigma^2}{\mu n}\cdot\frac{mn-\mu n}{mn-1}$$

podemos comparar as expressões (3. 1) e (3. 2) tal como em 2. relativamente à fórmula (2. 7).

Fazendo $\mu=1$ e calculando o limite quando n aumenta indefinidamente vê-se que

$$\bar{\rho} = \frac{\sigma_1^2}{\sigma^2}$$

e que serve, para alguns estatísticos, como definição de coeficiente de correlação.

Suponhamos que a característica estudada é quantitativa com duas alternativas; seja:

p_i — a proporção dos elementos do grupo i para os quais $x=1$;

$p=\frac{1}{m}\sum_{i=1}^{m} p_i$ — a proporção para o conjunto dos grupos.

Sabe-se, da análise de variância, que:

$$(3.3)\qquad pq = \frac{1}{m}\sum_{i=1}^{m} p_i q_i - \sigma_1^2$$

donde, atendendo a (2. 4), se tem:

$$(3.4)\qquad \bar{\rho} = \frac{1}{mpq}\left[\sum_{i=1}^{m}(p_i-p)^2 - \frac{1}{n-1}\sum_{i=1}^{m} p_i q_i\right]$$

que permite calcular a correlação interna.

4. Amostragem por etapas.

Consideremos um universo constituído por k unidades primárias (abreviadamente u. p.) das quais l são tiradas pelo método exaustivo. Cada u. p. compreende um número variável m_i $(1\leqq i\leqq l)$ de unidades secundárias. De cada u. p. tiraremos pelo processo elementar n_i unidades secundárias (abreviadamente u. s.), para formar a amostra pretendida que terá o efectivo

$$(4.1)\qquad n=\sum_{i=1}^{l} n_i .$$

Seja x_i^j o valor da característica X a estudar, na u. s. número j da u. p. número i.

Vamos ver que a grandeza

$$(4.2)\qquad \mu^* = \frac{1}{l}\sum_{i=1}^{l}\frac{m_i}{n_i}\sum_{j=1}^{n_i} x_i^j$$

dá um estimador centrado de

$$(4.3)\qquad \mu = \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{m_i} x_i^j$$

isto é

$$E[\mu^*]=\mu$$

Definam-se as grandezas $\bar{x}_i$, $\bar{\bar{x}}_i$, $\hat{x}_i$ pelas expressões

$$(4.4)\qquad \bar{x}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} x_i^j$$

$$(4.5)\qquad \bar{\bar{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_i^j$$

$$(4.6)\qquad \hat{x}_i = \sum_{j=1}^{m_i} x_i^j .$$

Considerando a u. p. número i como um universo distinto tem-se

$$(4.7)\qquad E_i[\mu^*] = \frac{1}{l}\sum_{i=1}^{l}\hat{x}_i$$

e, de acordo com (4. 6), vem, então

$$(4.8) \quad E[\mu^*] = E[\hat{x}_i] = \frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{m_i} x_i^j$$

como se pretendia.

Para calcular a variância de μ^* notemos que

$$(4.9) \quad E[(m_i \overline{\overline{x}}_i - \mu)^2] = E[m_i^2 (\overline{\overline{x}}_i - \overline{x}_i)^2] + + E[(\hat{x}_i - \mu)^2]$$

$$(4.10) \quad v_i[\overline{\overline{x}}_i] = \frac{\sigma_i^2}{n_i} \cdot \frac{m_i - n_i}{m_i - 1}$$

$$(4.11) \quad v[\hat{x}_i] = \sigma^2 \cdot \frac{k - l}{k - 1}$$

onde σ_i^2 é a variância dos x_i^j no interior da u. p. número i. Tem-se pois

$$(4.12) \quad v[\mu^*] = \frac{1}{l^2} \left\{ \sum_{i=1}^{l} E[(m_i \overline{\overline{x}}_i - \mu)^2] \right\} = = \frac{1}{l} \left[\sigma^2 \frac{k-l}{k-1} + \frac{1}{l} \sum_{i=1}^{l} m_i^2 \frac{\sigma_i^2}{n_i} \cdot \frac{m_i - n_i}{m_i - 1} \right]$$

onde a primeira parcela do último membro representa a parte de $v[\mu^*]$ devida às flutuações da amostra constituída pelas u. p. e costuma chamar-se-lhe *variância entre as unidades primárias de amostragem.*

Quanto à segunda parcela, ela representa a parte da variância imputável às flutuações da amostra constituída pelas u. s. no interior da u. p. número i e chama-se-lhe *variância entre as unidades secundárias no interior das unidades primárias de amostragem.*

A generalização a mais etapas torna-se complicada pois mesmo para o caso de três tipos de unidades seria necessário considerar em (4. 12) mais um termo que traduzisse a *variância entre as unidades terciárias no interior das unidades secundárias* (1).

(1) Para um estudo mais detalhado veja-se [3].

5. Amostragem por estratificação.

Consideremos uma população de N unidades dividida prèviamente em l subpopulações — estratos — de $N_1, \cdots, N_l$ unidades respectivamente, e de tal forma que

$$(5.1) \quad \sum_{h=1}^{l} N_h = N$$

após o que se tira independentemente uma amostra de cada estrato podendo ou não ter o mesmo efectivo.

A amostra estratificada será constituída por

$$(5.2) \quad n = \sum_{h=1}^{l} n_h$$

elementos; n_h é o número de unidades da amostra colhida no estrato h.

Designaremos por

x_{hi} — valor da unidade número i no estrato h;

$\overline{X}$ — média verdadeira da população;

$\overline{X}_h$ — média verdadeira no estrato h;

S_h^2 — variância verdadeira do estrato h;

$\overline{x}$; $\overline{x}_h$, s_h^2 representarão os valores observados, para a população e estrato.

Usando o estimador

$$(5.3) \quad \overline{x}_{st} = \frac{\sum_{h=1}^{l} N_h \overline{x}_h}{N}$$

conclui-se

$$(5.4) \quad \overline{x}_{st} = \overline{x} \Longleftrightarrow \frac{N}{n} = \frac{N_h}{n_h} = \text{constante.}$$

Diz-se que a estratificação é *proporcional.*

As principais propriedades de $\overline{x}_{st}$ aplicáveis a amostras estratificadas são as que se descrevem nos seguintes teoremas:

TEOREMA 1. *Se em cada estrato $\bar{x}_h$ é um estimador centrado de $\bar{X}_h$, então $\bar{x}_{st}$ é um estimador centrado da média $\bar{X}$.*

Seja

$$(5.5) \qquad \sigma^2(\bar{x}_h) = E[(\bar{x}_h - \bar{X}_h)^2]$$

e admitamos que

a) $E[\bar{x}_h] = \bar{X}_h$;

b) as amostras são extraídas independentemente nos diferentes estratos.

Nestas condições tem-se:

TEOREMA 2. *Para a amostra estratificada a variância de $\bar{x}_{st}$, como estimador de $\bar{X}$ é*

$$(5.6) \qquad \sigma^2(\bar{x}_{st}) = \frac{\sum\limits_{h=1}^{l} N_h^2 \sigma^2(\bar{x}_h)}{N^2}.$$

Este teorema diz-nos que $\sigma^2(\bar{x}_{st})$ depende sòmente das variâncias das médias individuais de cada estrato.

Como em cada estrato individualmente se tem

$$\sigma^2(\bar{x}_h) = \frac{S_h^2}{n_h} \cdot \frac{N_h - n_h}{N_h}$$

segue-se

$$(5.6') \qquad \sigma^2(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{l} N_h(N_h - n_h) \frac{S_h^2}{n_h}.$$

Em particular se as fracções de amostra

$$(5.7) \qquad f_h = \frac{n_h}{N_h}$$

são desprezáveis em todos os estratos virá

$$(5.8) \qquad \sigma^2(\bar{x}_{st}) = \sum_{h=1}^{l} \frac{W_h^2 S_h^2}{n_h}$$

onde W_h é o peso do estrato h.

Se admitirmos ainda que a estratificação é proporcional, a variância reduz-se a

$$(5.9) \qquad \sigma^2(\bar{x}_{st}) = \frac{N-n}{nN} \sum_{h=1}^{l} W_h S_h^2.$$

O uso da estratificação envolve as seguintes operações:

1) escolha da variável de estratificação;
2) escolha do número l de estratos;
3) determinação do processo de estratificação;
4) escolha do efectivo da amostra em cada estrato.

O teorema seguinte indica-nos um resultado importante estabelecido por TSCHUPROW (1923) e NEYMANN (1934).

TEOREMA 3. *Numa amostra aleatória estratificada, a variância $\sigma^2(\bar{x}_{st})$ é mínima, para um total de amostra fixo, se a amostra é repartida com n_h proporcional a $N_h S_h$.*

A demonstração pode fazer-se minimizando (5.6') sujeita a (5.2); com efeito, da relação

$$(5.10) \qquad F = \sigma^2(\bar{x}_{st}) + \lambda \left(\sum_{h=1}^{l} n_h - n \right)$$

obtem-se

$$(5.11) \qquad n_h = \frac{N_h S_h}{N\sqrt{\lambda}} = n \frac{N_h S_h}{\sum\limits_{h=1}^{l} N_h S_h}$$

valor que substituído em $\sigma^2(\bar{x}_{st})$ dá

$$(5.12) \qquad \sigma^2_{\min}(\bar{x}_{st}) = \frac{1}{N^2} \left[\frac{\left(\sum\limits_{h=1}^{l} N_h S_h \right)^2}{n} - \sum_{h=1}^{l} N_h S_h^2 \right].$$

Admitindo que a função custo se compõe de uma parte directamente proporcional ao volume da amostra (dentro de um estrato), mas onde o custo por unidade varia de estrato para estrato, e de outra parte visando gastos diversos podemos escrever

$$(5.13) \qquad C = a + \sum_{h=1}^{l} C_h n_h .$$

Ter-se-á neste caso

TEOREMA 4. *Com uma função custo do tipo* (5.13) *a variância* $\sigma^2(\bar{x}_{st})$ *é mínima quando*

$$(5.14) \qquad n_h = n \cdot \frac{\dfrac{N_h S_h}{\sqrt{C_h}}}{\sum\limits_{h=1}^{l} \dfrac{N_h S_h}{\sqrt{C_h}}},$$

o que conduz às seguintes regras: num dado estrato toma-se uma maior amostra se

a) *o estrato é maior;*

b) *o estrato é mais heterogéneo;*

c) *o custo da operação é mais barato.*

Vamos fazer seguidamente algumas considerações que permitirão estudar um critério (Teorema 5) de comparação das precisões obtidas pelos métodos da repartição óptima, proporcional e aleatória simples.

Do estudo da decomposição de quadrados e porque $\dfrac{1}{N_h}$ é desprezável pode escrever-se

$$(5.15) \qquad N S^2 = \sum_{h=1}^{l} [N_h S_h^2 + N_h (\bar{X}_h - \bar{X})^2].$$

Quando n é muito pequeno quando comparado com N tem-se

$$(5.16) \qquad \sigma_a^2 = \frac{S^2}{n}$$

$$(5.17) \qquad \sigma_{prop}^2 = \frac{\sum\limits_{h=1}^{l} N_h S_h^2}{n N}$$

$$(5.18) \qquad \sigma_{opt}^2 = \frac{\left(\sum\limits_{h=1}^{l} N_h S_h\right)^2}{n N^2}$$

e portanto

$$(5.19) \qquad \sigma_a^2 = \sigma_{prop}^2 + \frac{\sum\limits_{h=1}^{l} N_h (\bar{X}_h - \bar{X})^2}{n N}$$

Da definição de σ_{opt}^2 resulta

$$(5.20) \qquad \sigma_{prop}^2 - \sigma_{opt}^2 =$$

$$= \frac{1}{n N} \left[\sum_{h=1}^{l} N_h S_h^2 - \frac{\left(\sum\limits_{h=1}^{l} N_h S_h\right)^2}{N} \right].$$

Definindo $\bar{S}$ pela igualdade

$$(5.21) \qquad \bar{S} = \frac{\sum\limits_{h=1}^{l} N_h S_h}{N}$$

virá, atendendo a (5.19) e (5.20)

$$(5.22) \qquad \sigma_a^2 = \sigma_{opt}^2 + \frac{\sum\limits_{h=1}^{l} N_h (S_h - \bar{S})^2}{n N} +$$

$$+ \frac{\sum\limits_{h=1}^{l} N_h (\bar{X}_h - \bar{X})^2}{n N}.$$

Nas mesmas condições do teorema pode estabelecer-se

$$(5.23)\qquad \sigma_a^2 = \sigma_{\text{prop}}^2 + \frac{(N-n)}{Nn(N-1)} \cdot$$

$$\cdot \left[\sum_{h=1}^{l} N_h (\overline{X}_h - \overline{X})^2 - \frac{1}{N} \sum_{h=1}^{l} (N - N_h) S_h^2 \right].$$

Ter-se-á portanto:

TEOREMA 5. *Se os termos* f_h *são desprezáveis tem-se*

$$(5.24)\qquad \sigma_{\text{opt}}^2 \leqq \sigma_{\text{prop}}^2 \leqq \sigma_a^2$$

em que a repartição óptima é para um n *fixo e* n_h *dado por* (5.11).

Suponhamos que a população em estudo pode ser representada pela função de frequência $f(x)$ $(a \leqq x \leqq b)$ e que pretendemos dividi-la em l estratos.

Considerem-se as grandezas W_h, μ, μ_h definidas por

$$(5.25)\qquad W_h = \int_{x_{h-1}}^{x_h} f(\alpha)\, d\alpha$$

$$(5.26)\qquad \mu = \sum_{h=1}^{l} W_h \cdot \mu_h$$

$$(5.27)\qquad \mu_h = \frac{1}{W_h} \int_{x_{h-1}}^{x_h} \alpha \cdot f(\alpha)\, d\alpha .$$

Admitamos ainda que a função custo se obtém de (5.13) fazendo $a = 0$ para qualquer dos quatro casos seguintes:

1) repartição proporcional

$$n_h \approx n \cdot W_h$$

2) repartição com variância mínima

$$n_h \approx S_h \cdot W_h$$

3) repartição óptima

$$n_h \approx \frac{S_h \cdot W_h}{\sqrt{C_h}}$$

4) especial $n_1 < N_1$, $n_2 = N_2$

a que correspondem as variâncias

$$(5.28)\qquad \sigma_{\text{prop}}^2 = \frac{\sum_{h=1}^{l} W_h S_h^2}{n}$$

$$(5.29)\qquad \sigma_{\text{min}}^2 = \frac{\left(\sum_{h=1}^{l} W_h S_h\right)^2}{n}$$

$$(5.30)\qquad \sigma_{\text{opt}}^2 = \frac{\left(\sum_{h=1}^{l} W_h S_h \sqrt{C_h}\right)^2}{C}$$

$$(5.31)\qquad \sigma_{\text{esp}}^2 = W_1^2 \, \frac{S_1^2}{n_1} \cdot \frac{N_1 - n_1}{N_1 - 1}.$$

Definindo a função

$$\Phi_\bullet(x_1, \cdots, x_{l-1}) = n \cdot \sigma_\bullet^2$$

deduzem-se os pontos divisórios x_h a partir da igualdade

$$\frac{\partial \Phi_\bullet}{\partial x_h} = 0$$

Quanto à determinação do número óptimo de estratos a ideia básica é a possibilidade de diminuição da variância de um estimador aumentando o número de estratos. Assim, para um estimador $\hat{\theta}$, e para cada valor de

l, computam-se os valores $\sigma^2(\hat{\theta}\,;l)$ e pode dizer-se que

$$G = \sigma^2(\hat{\theta}\,;l-1) - \sigma^2(\hat{\theta}\,;l)$$

representa o ganho, em precisão, devido ao aumento do número de estratos.

A questão delicada é determinar uma relação que traduza, em cada caso concreto, o modo como variam

$$\sigma^2(\hat{\theta}\,;1) \text{ e } \sigma^2(\hat{\theta}\,;l) \qquad l > 1$$

admitindo em tudo o que se disse que o estimador é linear.

*

* *

Este trabalho pretende dar algumas indicações muito gerais sobre certos problemas da teoria da amostragem, mostrando as vantagens e inconvenientes que aparecem nos diferentes tipos de amostragem considerados.

Por falta de tempo não foi possível fazer o estudo da determinação do número óptimo de estratos além dos casos analisados por Dalenius [2] com as funções de frequência $f(x) = 1$, $f(x) = e^{-x}$ e $f(x) = x\,e^{-x}$; para as duas últimas sugeriu uma relação do tipo

$$\sigma^2(\hat{\theta}\,;l) = \frac{1}{\left(\dfrac{l}{1-l}\right)^2}\,\sigma^2(\hat{\theta}\,;l-1)$$

enquanto para a primeira obteve

$$\sigma^2(\hat{\theta}\,;k\,l) = \frac{1}{k^2}\,\sigma^2(\hat{\theta}\,;l)$$

Também era intenção testar a tabela [5] mas como se pretende ir até aos 10.000 números, só então o faremos.

048353	814391	621032	996691
050464	431130	694510	815310
116654	715461	547388	606533
154708	578817	866132	267673
227814	635385	020554	363950
242031	504103	602206	114080
256814	046143	275966	046249
329401	775413	081605	730905
332235	498178	167277	733028
402155	065299	003607	685231
410801	540368	866636	376959
536636	981401	508121	367880
548182	254898	677122	920347
592369	620259	551168	572190
666488	590136	645844	904475
680845	051960	206773	091913
717442	914135	248036	200358
735340	031841	184284	406448
762768	060569	257898	373732
774669	213371	833357	178320
808781	126376	752491	504864
846617	957137	032188	528942
896886	371617	655202	201562
947131	213648	502064	326807
957141	168828	770483	064670

Extraída de [5].

BIBLIOGRAFIA

[1] Cochram, William *Sampling Techniques* — 1963 — 2.ª edição — John Wiley & Sons, Inc. New York.

[2] Dalenius, Tore *Sampling on Sweden* — 1957 — Almqvst & Wiksell — Estocolmo.

[3] *La Theorie des sondages* N.os 5 e 6 — 1953 — I. N. S. E. E. Paris.

[4] Soares, Rui *Alguns aspectos da Teoria da amostragem* — 1969, Faculdade de Ciências de Lisboa.

[5] ———, Tábua de 5.000 números aleatórios.