Amostragem por Grupos

Na amostragem por grupos as **unidades estatísticas são agrupadas**, de acordo com algum critério. É habitual designar cada grupo por **unidade primária** sendo os elementos que compõem cada unidade primária designados por **unidades secundárias.**

Cada grupo ou unidade primária é seleccionado aleatoriamente e os valores de todas as unidades secundárias que o compõem são observados.

Note-se que, enquanto na amostragem estratificada há uma escolha aleatória das unidades estatísticas a observar em cada estrato, aqui observam-se (regra geral) todas as unidades que compõem o grupo. A escolha aleatória é feita na selecção dos grupos.

Este tipo de amostragem é usado quando não se conhece bem a população ou por questão de recursos, as áreas onde se vai investigar são limitadas.

Enquanto as razões que levam à estratificação é a sua potencialidade de obter estimadores mais eficientes para as características da população, a amostragem por grupos é muitas vezes justificada por razões de conveniência ou porque não se dispõe de uma base de amostragem completa, e portanto para melhorar a possibilidade de acesso à população ou ainda por razões de custo.

Exemplo

No estudo de uma dada doença, não se possui uma lista completa das pessoas que dela padecem, então opta-se pela selecção de unidades hospitalares e escolhida uma delas observam-se todos os portadores da referida doença aí existentes.

Fazendo mais uma vez comparação entre a amostragem estratificada e amostragem por grupos, na primeira os estratos devem ser homogéneos, na segunda obtém-se maior precisão quanto maior for a heterogeneidade interna dos grupos, no que respeita às características que a partir deles se pretendam estudar.

Notação:

M- nº total de unidades (primárias) em que a população foi dividida

m- nº de unidades primárias incluídas na amostra

N_i – nº de unidades secundárias pertencentes à unidade primária i

$N = \sum_{i=1}^{M} N_i$ - nº total de unidades secundárias na população

$f = \frac{m}{M}$ - fracção de amostragem

x_{ij} - valor associado à j.ésima unidade secundária da unidade primária i.

Neste tipo de amostragem, como dissemos atrás os m grupos a incluir na amostra são seleccionados aleatoriamente.

A-- Grupos de igual dimensão

Neste caso

$N_1 = N_2 = ... = N_M = L$

então $N = ML$ e a observação de m grupos dá $n = mL$ observações.

A fracção de amostragem é

$$f = \frac{n}{N} = \frac{mL}{ML} = \frac{m}{M}$$

Sejam então as observações recolhidas

x_{ij} $\quad i = 1,...,m \quad j = 1,...,L$ $\quad \mu_i \quad \sigma_i'^2$ designam a média e a variância do grupo

Estimação do valor médio da população

$$\bar{x}_{cl} = \frac{1}{mL}\sum_{i=1}^{m}\sum_{j=1}^{L} x_{ij} = \frac{1}{m}\sum_{i=1}^{m}\mu_i$$

$$E\left[\bar{X}_{cl}\right] = \mu$$

$$Var\left[\bar{X}_{cl}\right] = \frac{1}{m}(1-f)\sum_{i=1}^{M}\frac{(\mu_i - \mu)^2}{M-1}$$

que é estimada por

$$S^2\left[\bar{X}_{cl}\right] = \frac{1}{m(m-1)}(1-f)\sum_{i=1}^{m}\left(\mu_i - \bar{x}_{cl}\right)^2$$

B-- Amostragem em grupos com diferentes tamanhos

M grupos

mas $N_1, N_2, ..., N_M$ não têm todos o mesmo valor

$$N = \sum_{i=1}^{M} N_i$$

Sendo assim a dimensão da amostra é agora uma **variável aleatória** porque o seu valor depende dos grupos seleccionados para a amostra.

$$n = \sum_{i=1}^{m} N_i$$

Agora o tratamento mais complicado

Selecção com probabilidades iguais sem reposição

Neste caso a média da população é

$$\mu = \frac{1}{N} \sum_{i=1}^{M} \sum_{j=1}^{Ni} x_{ij} = \sum_{i=1}^{M} \frac{Ni}{N} \mu_i \qquad \text{com} \qquad \mu_i = \sum_{i=1}^{Ni} \frac{xij}{Ni}$$

definindo $\overline{N} = \frac{N}{M}$ pode escrever-se

$$\mu = \frac{1}{M} \sum_{i=1}^{M} \frac{Ni}{\overline{N}} \mu_i$$

Estimação do valor médio e do total

A média da população é estimada por

$$\overline{x}_{cl} = \frac{1}{m} \sum_{i=1}^{m} \frac{Ni}{\overline{N}} \mu_i$$

$$\hat{\mathrm{var}}(\overline{x}_{cl}) = \frac{1}{m(m-1)} (1-f) \sum_{i=1}^{m} \left(\frac{Ni}{\overline{N}} \mu_i - \overline{x}_{cl} \right)^2 = \frac{1}{m(m-1)} (1-f) \sum_{i=1}^{m} \left(\frac{(N_i \mu_i)^2}{\overline{N}^2} - m\, \overline{x}_{cl}^2 \right)$$

O total da população é estimado por

$$\hat{T}_{cl} = \frac{M}{m}\sum_{i=1}^{m} N_i \mu_i$$

$$\widehat{\mathrm{var}}\left(\hat{T}_{cl}\right) = \frac{1}{m(m-1)}(1-f)\sum_{i=1}^{m}\left(MN_i\mu_i - \hat{T}_{cl}\right)^2$$