MAT02036 - Amostragem 2

Aula 12 - Amostragem por Conglomerados - Parâmetros e Estimação

Markus Stein

Departamento de Estatística, IME/UFRGS

2022/2

Housekeeping

- Aproveitem o momento presencial para tirar dúvidas
- Se estivéssemos no ensino remoto ou à distância
  - vocês poderiam estar somente ouvindo, sem interação
  - ou assistindo vídeos e material em outro momento
- Depois das aulas, rever material da aula passada
  - fazer exercícios
  - se preparar para a próxima aula

Aula passada 💾

Amostragem por Conglomerados (*Cluster*)

A população de unidades U é particionada em M grupos **mutuamente exclusivos e exaustivos**,

$$U = C_1 \cup C_2 \cup \cdots \cup C_M = \bigcup_{i=1}^{M} C_i \quad \text{e} \quad C_i \cap C_k = \emptyset, i \neq k$$

As unidades são identificadas por dois índices i, j, onde i designa o conglomerado e j designa o rótulo da unidade dentro do conglomerado,

$$C_i = \{(i,1), (i,2), \ldots, (i,j), \ldots, (i, N_i)\}, \text{ para } i = 1, 2, \ldots, M.$$

N_i é o tamanho do conglomerado C_i. Então $N = N_1 + N_2 + \cdots + N_M$ é o tamanho total da população.

Selecione uma amostra $a = \{i_1, \ldots, i_m\}$ de tamanho $m \ (m > 0)$, entre os rótulos de $C = \{1, \ldots, M\}$ para selecionar os conglomerados, segundo um plano amostral $p(a)$.

Aula passada

Amostragem por conglomerados em um estágio

- Num plano de **Amostragem por Conglomerados em 1 estágio (AC1)**, todas as unidades populacionais dos conglomerados selecionados em $a = (i_1, \ldots, i_m)$ farão parte da amostra,

$$s = C_{i_1} \cup C_{i_2} \cup \cdots \cup C_{i_m} = \bigcup_{k=1}^{m} C_{i_k}$$

Portanto, a amostragem do tipo **AC1** é caracterizada pelos seguintes aspectos:

- As unidades populacionais são reunidas em *grupos* denominados *conglomerados*.

- Uma amostra de unidades é obtida selecionando uma *amostra de conglomerados* e *incluindo na amostra todas as unidades* pertencentes aos conglomerados selecionados.

Aula passada

Amostragem conglomerada em vários estágios

- Numa **amostragem conglomerada em três estágios**:
  - **Unidades Primárias** de Amostragem **- UPA**s;
  - **Unidades Secundárias** de Amostragem **- USA**s;
  - e **unidades elementares**.

Aula passada

Notação na amostragem por conglomerados em um estágio

Os **tamanhos populacionais** e da **amostra** na **AC1** para uma dada população são

Definição	População	Amostra
Conglomerados	M	m
Unidades no conglomerado i	N_i	N_i
Tamanho da população/amostra	$N = \sum_{i \in C} N_i$	$n = \sum\limits_{i \in a} N_i$

- $C = \{1, \ldots, M\}$: índice dos conglomerados na população.
- $a = (i_1, \ldots, i_m)$: índice dos conglomerados selecionados para a amostra.

Aula passada

Notação na amostragem por conglomerados em um estágio

Os parâmetros populacionais **total** e **média** por conglomerado e na população como um todo são dados por

Definição	Parâmetro
Valor da variável de pesquisa para unidade j do conglomerado i	y_{ij}
Total no conglomerado i	$T_i = \sum_{j \in C_i} y_{ij}$
Média no conglomerado i	$\overline{Y_i} = T_i/N_i = \frac{1}{N}_i \sum_{j \in C_i} y_{ij}$
Total populacional	$T = \sum_{i=1}^{M} T_i = \sum_{i \in C} T_i$
Média populacional por conglomerado	$\overline{Y_C} = T/M = \frac{1}{M} \sum_{i \in C} T_i$
Média populacional por unidade	$\overline{Y} = T/N = \frac{1}{N} \sum_{i \in C} T_i$

para $i = 1, \ldots, M$ e $j = 1, \ldots, N_i$.

Amostragem por Conglomerados em 1 Estágio Simples

Amostragem por Conglomerados em 1 Estágio Simples

Parâmetros

Exercício

Mostrar que

$$\overline{Y} = \overline{Y}_C$$

se $N_i = \overline{N}$ (conglomerados de mesmo tamanho), $\overline{N} = \frac{N}{M}$, para todo $i = 1, \ldots, M$.

Amostragem por Conglomerados em 1 Estágio Simples

Variâncias da população geral e dos conglomerados:

- A **variância** das unidades **no mesmo conglomerado** i é dada por

$$Var_{i,y} = \frac{\sum_{j=1}^{N_i} \left(y_{ij} - \overline{Y}_i \right)^2}{N_i} \text{ ou } S_{i,y}^2 = \frac{\sum_{j=1}^{N_i} \left(y_{ij} - \overline{Y}_i \right)^2}{N_i - 1}$$

- A **variância global** das unidades é dada por

$$Var_y = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N_i} \left(y_{ij} - \overline{Y} \right)^2}{N} \text{ ou } S_y^2 = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N_i} \left(y_{ij} - \overline{Y} \right)^2}{N - 1}$$

Amostragem por Conglomerados em 1 Estágio Simples

Variâncias da população geral e dos conglomerados:

Podemos definir:

- A **variância dentro** dos conglomerados

$$Var_{dc} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N_i}\left(y_{ij} - \overline{Y}_i\right)^2}{N} = \frac{1}{\overline{N}M}\sum_{i=1}^{M}\frac{N_i}{N_i}\sum_{j=1}^{N_i}\left(y_{ij} - \overline{Y}_i\right)^2 = \frac{1}{M}\sum_{i=1}^{M}\frac{N_i}{\overline{N}}\sum_{j=1}^{N_i} Var_{i,y}.$$

- A **variância entre** os conglomerados

$$Var_{ec} = \frac{\sum_{i=1}^{M} N_i\left(\overline{Y}_i - \overline{Y}\right)^2}{N}$$

Exercício

Mostre: (❓)

$$Var_y = Var_{dc} + Var_{ec}.$$

Amostragem por Conglomerados em 1 Estágio Simples

Variâncias da população geral e dos conglomerados:

- Ou ainda,

$$S_y^2 = \frac{(\overline{N} - 1)MS_{dc}^2 + \overline{N}(M-1)\overline{S}_{ec}^2}{M\overline{N} - 1},$$ (❓)

onde $\overline{S}_{ec}^2 = \frac{S_{ec}^2}{\overline{N}}$ e:

- a medida da **variância entre** os totais dos conglomerados

$$S_{ec}^2 = \frac{1}{M-1} \sum_{i \in C} \left(T_i - \overline{Y}_C \right)^2;$$

- a medida da variância *dentro* dos conglomerados, dada por:

$$S_{dc}^2 = \frac{1}{M} \sum_{i \in C} \frac{1}{\overline{N} - 1} \sum_{j \in C_i} \left(y_{ij} - \overline{Y_i} \right)^2 = \frac{1}{M} \sum_{i \in C} S_i^2.$$

Amostragem por Conglomerados em 1 Estágio Simples

Exercício

A seguir está um quadro com a população de lojas de um bairro de um município agrupadas em conglomerados (quarteirões). A variável a ser estudada é o número de funcionários dessas lojas. Calcule os parâmetros tamanho médio dos conglomerados, total populacional, média populacional e por conglomerado e a média das médias.

Núm. quarteirão	Núm. de lojas	Núm. de funcionários
1	4	12
2	6	24
3	2	10
4	6	12
Total	18	58

Amostragem por Conglomerados em 1 Estágio Simples

Plano amostral e probabilidades de inclusão

1. Na Amostragem por Conglomerados em 1 Estágio Simples (**AC1S**) sorteamos $m < M$ conglomerados através de **AAS** (com ou sem reposição).

2. De cada conglomerado são observados todas as unidades. Assim, podemos pensar na **AC1S** como uma **AAS** de valores agregados dos conglomerados:

$$U_C = \{C_1, C_2, \ldots, C_M\}$$

$$\boldsymbol{T} = (T_1, T_2, \ldots, T_M)\,.$$

Exemplo: Seja uma AC1S de $m = 2$.

A probabilidade de um conglomerado C_1 pertencer à amostra é:

Amostragem por Conglomerados em 1 Estágio Simples

Plano amostral e probabilidades de inclusão

- NA **AC1S sem reposição**

$$
\begin{aligned}
P(C_1 \in s) &= P\left(\text{``}C_1 \text{ na } 1^a \text{ extração ''} \; \cup \text{``}C_1 \text{ na } 2^a \text{ extração ''}\right) \\
&= P\left(\text{``}C_1 \text{ na } 1^a \text{ extração ''}\right) + P\left(\text{``}C_1 \text{ na } 2^a \text{ extração ''}\right) \\
&= \frac{1}{M} + P\left(\text{``}\overline{C}_1 \text{ na } 1^a \text{ extração ''}\right) P\left(\text{``}C_1 \text{ na } 2^a \text{ extração ''} \middle| \text{``}\overline{C}_1 \text{ na } 1^a \text{ extração ''}\right) \\
&= \frac{1}{M} + \frac{M-1}{M}\frac{1}{M-1} = \frac{2}{M}.
\end{aligned}
$$

- NA **AC1S com reposição**

$$
P(C_1 \in s) = 1 - P(C_1 \notin s) =
$$

$$
1 - P\left(\text{``}\overline{C}_1 \text{ na } 1^a \text{ extração ''} \; \cap \text{``}\overline{C}_1 \text{ na } 2^a \text{ extração ''}\right)
$$

$$
1 - \left(1 - \frac{1}{M}\right)\left(1 - \frac{1}{M}\right) = 1 - \left(1 - \frac{1}{M}\right)^2.
$$

Amostragem por Conglomerados em 1 Estágio Simples

Plano amostral e probabilidades de inclusão

- Na **AC1**, as **probabilidades de inclusão de um conglomerado** também são as **probabilidades de inclusão de um especifico elemento** da população (i,j) na amostra já que:

$$P\left[(i,j) \in s\right] = P(C_i \in s).$$

- Assim, **todos os elementos têm a mesma probabilidade** de serem incluídos na amostra, independente do tamanho do conglomerado a que pertencem.

Amostragem por Conglomerados em 1 Estágio Simples

Informações amostrais nos conglomerados selecionados

Descrição	Valores amostrais
Valor da variável de pesquisa para unidade j do conglomerado selecionado i	$y_{ij} \forall j = 1, \ldots, N_i$, $i \in a = \{i_1, \ldots, i_m\}$
Total no conglomerado i da amostra $i \in a = \{i_1, \ldots, i_m\}$	$T_i = \sum_{j \in C_i} y_{ij}$
Média no conglomerado i da amostra $i \in a = \{i_1, \ldots, i_m\}$	$\overline{Y_i} = T_i / N_i$
Total amostral	$t = \sum_{i \in a} T_i$
Média por conglomerado	$\bar{y}_C = t/m = \frac{1}{m} \sum_{i \in a} T_i$
Média por unidade	$\bar{y} = t/n = \sum_{i \in a} T_i \Big/ \sum_{i \in a} N_i$

Amostragem por Conglomerados em 1 Estágio Simples

Estimação do total sob AC1S - estimador natural

- O **estimador** 'natural' (de Horvitz-Thompson) do **total** populacional sob plano amostral **AC1S** é dado por:

$$\widehat{T}_{AC1S/HT} = \frac{M}{m} \sum_{i \in a} T_i = M \bar{y}_C = \sum_{i \in a} \sum_{j \in C_i} d_{ij} y_{ij}$$

onde $d_{ij} = M/m$ são os pesos individuais básicos sob **AC1S**.

- $\widehat{T}_{AC1S/HT}$ é um estimador não viciado para o total populacional T.
  - Sabemos mostrar $E\left(\widehat{T}_{AC1S/HT}\right) =$ ❓.

Amostragem por Conglomerados em 1 Estágio Simples

Estimação do total sob AC1S - estimador natural

A **variância do estimador** natural do **total** populacional é dada por:

- Na **AC1S SEM reposião**

$$Var_{AC1S}\left(\widehat{T}_{AC1S/HT}\right) = M^2\,(1-f)\,\frac{S_{ec}^2}{m} = M^2\left(\frac{1}{m}-\frac{1}{M}\right)S_{ec}^2$$

onde $f = m/M$ e $S_{ec}^2 = \frac{1}{M-1}\sum_{i\in C}\left(T_i - \overline{Y}_C\right)^2$ é a **variância entre** os totais dos conglomerados.

- Na **AC1S COM reposião**

$$Var_{AC1S}\left(\widehat{T}_{AC1S/HT}\right) = M^2\,\frac{Var_{ec_T}^2}{m}$$

em que nese caso $Var_{ec_T}^2 = \frac{1}{M}\sum_{i\in C}\left(T_i - \overline{Y}_C\right)^2$ é a **variância entre** os totais dos conglomerados.

Amostragem por Conglomerados em 1 Estágio Simples

Estimação do total sob AC1S - estimador natural

- Um **estimador não viciado** para a **variância do estimador** natural do total populacional é dado por:
  - **SEM reposição** de conglomerados, $\widehat{V}_{AC1S}\left(\widehat{T}_{AC1S/HT}\right) = M^2\left(1-f\right)\frac{\widehat{S}^2_{ec}}{m}$

onde $\widehat{S}^2_{ec} = \frac{1}{m-1}\sum_{i\in a}\left(T_i - \overline{y}_C\right)^2$.

- **COM reposição** de conglomerados, $\widehat{V}_{AC1S}\left(\widehat{T}_{AC1S/HT}\right) = M^2\frac{\widehat{S}^2_{ec}}{m}$

- Note que $\widehat{S}^2_{ec}$ é estimador não viesado de
  - $S^2_{ec} = \frac{M}{M-1}Var_{ec_T}$ se **AC1s SEM reposição** e
  - Var_{ec_T} se amostragem **AC1S COM reposição**.

Amostragem por Conglomerados em 1 Estágio Simples

Estimação do total sob AC1S - estimador de razão

- Quando todos os conglomerados têm tamanhos iguais, $N_i = \overline{N}, \forall i \in C$, o estimador natural é a única opção de estimador simples para o total.
  - Quando os tamanhos dos conglomerados variam, este estimador pode ser pouco eficiente.
- Um estimador que reduz o efeito da variação de N_i na precisão é um **estimador tipo razão**, baseado no tamanho dos conglomerados.

$$\widehat{T}^{R}_{AC1S} = \frac{N}{n} \sum_{i \in a} T_i = N\overline{y} = \sum_{i \in a} \sum_{j \in C_i} w^R_{ij} y_{ij}$$

  onde $w^R_{ij} = N/n$ são pesos amostrais ajustados ou 'calibrados', no sentido de que seu uso aplicado a uma variável de contagem $(y_{ij} = 1, \forall i, j)$ levaria a obter uma estimativa para o tamanho da população igual ao tamanho total N.

- Note que este estimador requer que o tamanho total da população N seja conhecido. Portanto, em muitas situações este estimador não é viável.

Amostragem por Conglomerados em 1 Estágio Simples

Estimação do total sob AC1S - estimador de razão

- A variância aproximada do estimador tipo razão do total, **SEM reposição** de conglomerados, é dada por:
    - $$Var_{AC1S}\left(\widehat{T}^{R}_{AC1S}\right) \doteq M^2\left(\frac{1}{m}-\frac{1}{M}\right)\frac{1}{M-1}\sum_{i\in C} N_i^2\left(\overline{Y}_i-\overline{Y}\right)^2.$$
- Esta **aproximação** requer que o **número de conglomerados** na amostra m seja **grande**.
- O **estimador da variância** do **estimador tipo razão** do **total**, **SEM reposição** de conglomerados, pode ser obtido por:

$$\widehat{Var}_{AC1S}\left(\widehat{T}^{R}_{AC1S}\right) = M^2\left(\frac{1}{m}-\frac{1}{M}\right)\frac{1}{m-1}\sum_{i\in a} N_i^2\left(\overline{Y}_i-\overline{y}\right)^2.$$

Amostragem por Conglomerados em 1 Estágio Simples

Estimação do total sob AC1S

Comparando os estimadores natural e tipo razão para o total populacional, sob o plano amostral AC1S, tem-se:

1. Se os conglomerados tiverem todos o mesmo tamanho, ou seja:

   $$N_i = N/M = \overline{N}, \forall i = 1, \ldots, M$$

   então:

   $$\widehat{T}^{R}_{AC1S} = \widehat{T}_{AC1S/HT}$$

2. Somente o estimador natural, $\widehat{T}_{AC1S/HT}$, pode ser utilizado quando N for desconhecido.

3. O estimador $\widehat{T}_{AC1S/HT}$ é exatamente **não viciado**.

4. O estimador $\widehat{T}^{R}_{AC1S}$ é apenas **aproximadamente não viciado**, para grandes amostras.

Amostragem por Conglomerados em 1 Estágio Simples

Estimação do total sob AC1S

1. O estimador tipo razão, $\widehat{T}^{R}_{AC1S}$, pode ser muito mais preciso que $\widehat{T}_{AC1S/HT}$ em certos casos, pois se:

$$\overline{Y_i} \doteq \overline{Y}, \forall i$$

então:

$$Var_{AC1S}\left(\widehat{T}^{R}_{AC1S}\right) \doteq 0;$$

enquanto que

$$Var_{AC1S}\left(\widehat{T}_{AC1S/HT}\right) \propto \sum_{i \in C}\left(T_i - \overline{Y}_C\right)^2 =$$

$$\sum_{i \in C}\left(N_i\overline{Y_i} - \overline{N}\,\overline{Y}\right)^2 \doteq \overline{Y}^2 \sum_{i \in C}\left(N_i - \overline{N}\right)^2.$$

Amostragem por Conglomerados em 1 Estágio Simples

Estimação do total sob AC1S

- Isto é, a variância do estimador natural incorpora parcela devida à variação dos tamanhos dos conglomerados e, portanto, a ocorrência de variabilidade nos tamanhos dos conglomerados causa acentuada perda de precisão nesse estimador sob amostragem conglomerada em um estágio simples. Essa perda será maior quando maior for a variabilidade dos tamanhos dos conglomerados.

- Na prática, as médias $\overline{Y}_i$ são menos variáveis entre conglomerados que os totais Y_i, e portanto:

$$Var_{AC1S}\left(\widehat{T}^{R}_{AC1S}\right) < Var_{AC1S}\left(\widehat{T}_{AC1S/HT}\right)$$

Os ganhos de precisão do estimador tipo razão podem ser grandes quando:

- For grande a variação dos tamanhos N_i.
- For pequena a variação entre as médias $\overline{Y_i}$ dos conglomerados.

Amostragem por Conglomerados em 1 Estágio Simples

Estimação do total sob AC1S

Na prática, a formação de conglomerados com tamanhos iguais para controlar a variação de tamanho na variância do estimador e, também, na variação do tamanho final da amostra, nem sempre é possível, sendo a ocorrência de conglomerados com tamanhos iguais pouco comum.

Assim, ao invés de tentar construir artificialmente conglomerados de tamanhos iguais, é possível manter os conglomerados com tamanhos desiguais e utilizar métodos de seleção de amostra e estimadores adequados na expectativa de redução da variância e de menor perda de precisão com o uso da amostragem conglomerada.

Os métodos usuais para reduzir o efeito da variabilidade dos tamanhos dos conglomerados são:

a) Selecionar os conglomerados com probabilidades proporcionais ao tamanho.
b) Estratificar os conglomerados, utilizando o tamanho como variável de estratificação.
c) Usar estimadores tipo razão, com a variável auxiliar sendo o tamanho do conglomerado.

Amostragem por Conglomerados em 1 Estágio Simples

Estimação do total sob AC1S

Recomendações:

- Em geral, prefira $\widehat{T}^{R}_{AC1S}$ a menos que N seja desconhecido.

- Se $\widehat{T}_{AC1S/HT}$ tiver que ser usado: **estratifique os conglomerados** por tamanho ou **use amostragem conglomerada com PPT**. Em termos de eficiência não parece haver vantagem nítida de qualquer das duas alternativas, sendo bastante semelhantes os resultados obtidos com ambas as técnicas em termos da precisão final das estimativas.

Amostragem por Conglomerados em 1 Estágio Simples

Estimação da média sob AC1S - estimador natural

- O **estimador** de Horvitz-Thompson da **média** por unidade, $\overline{\overline{Y}}$, sob plano amostral **AC1S** é dado por:

$$\overline{y}_{AC1S/HT} = \frac{\widehat{T}_{AC1S/HT}}{N} = \frac{M}{N}\frac{1}{m}\sum_{i \in a} T_i = \overline{y}_C / \overline{\overline{N}}.$$

- A **variância do estimador** HT da **média** é dada por:
  - **SEM reposição**,

$$Var_{AC1S}\left(\overline{y}_{AC1/HT}\right) = \frac{M^2}{N^2}\left(\frac{1}{m} - \frac{1}{M}\right) S_{ec}^2 = \frac{1}{\overline{\overline{N}}^2}\left(\frac{1}{m} - \frac{1}{M}\right) S_{ec}^2;$$

  - **COM reposição**,

$$Var_{AC1S}\left(\overline{y}_{AC1/HT}\right) = \frac{1}{\overline{\overline{N}}^2}\frac{Var_{ec_T}^2}{m}.$$

Amostragem por Conglomerados em 1 Estágio Simples

Estimação da média sob AC1S - estimador natural

- O **estimador da variância do estimador** HT da **média** é dado por:
- **SEM reposição**,

$$\widehat{V}ar_{AC1S}\left(\overline{y}_{AC1S/HT}\right) = \frac{M^2}{N^2}\left(\frac{1}{m} - \frac{1}{M}\right)\widehat{S}^2_{ec} = \frac{1}{\overline{N}^2}\left(\frac{1}{m} - \frac{1}{M}\right)\widehat{S}^2_{ec}$$

- **COM reposição**,

$$\widehat{V}ar_{AC1S}\left(\overline{y}_{AC1S/HT}\right) = \frac{M^2}{N^2}\frac{\widehat{S}^2_{ec}}{m} = \frac{1}{\overline{N}^2}\frac{\widehat{S}^2_{ec}}{m}$$

Amostragem por Conglomerados em 1 Estágio Simples

Estimação da média sob AC1S - estimador de razão

- Um estimador tipo razão da média por unidade $\overline{Y}$ sob o plano amostral **AC1S**, **sem reposição** de conlgomerados, é dado por:

$$\overline{y}^{R}_{AC1S} = \frac{\widehat{T}^{R}_{AC1S}}{N} = \frac{1}{n}\sum_{i\in a} T_i = \overline{y}_C / \overline{n} = \overline{y}$$

onde $\overline{n} = \frac{1}{m}\sum_{i\in a} N_i = \frac{n}{m}$.

- A variância aproximada do estimador de razão da média é dado por:

$$Var_{AC1S}\left(\overline{y}^{R}_{AC1S}\right) \doteq \frac{1}{\overline{N}^2}\left(\frac{1}{m} - \frac{1}{M}\right)\frac{1}{M-1}\sum_{i\in C} N_i^2\left(\overline{Y_i} - \overline{Y}\right)^2$$

- Esta aproximação é válida somente para amostras grandes, isto é, com m grande.

Amostragem por Conglomerados em 1 Estágio Simples

Estimação da média sob AC1S - estimador de razão

- O **estimador da variância** do **estimador tipo razão** da **média** é dado por:

$$\widehat{V}ar_{AC1S}\left(\overline{y}^{R}_{AC1S}\right) = \frac{1}{\overline{n}^2}\left(\frac{1}{m} - \frac{1}{M}\right)\frac{1}{m-1}\sum_{i \in a} N_i^2\left(\overline{Y_i} - \overline{y}\right)^2$$

com $\overline{n}$ em lugar de $\overline{N}$ quando este for desconhecido.

Notas:

1. Se N (ou $\overline{N}$) for desconhecido, só podemos usar $\overline{y}^{R}_{AC1S}$.
2. As comparações de vício e variância feitas para o caso dos estimadores de total seguem válidas para os estimadores da média.
3. Quase sempre é preferível usar $\overline{y}^{R}_{AC1S} = \overline{y}$, a média simples por unidade elementar.

Amostragem por Conglomerados em 1 Estágio Simples

Estimação da proporção sob AC1S

- Ainda não foi tratado explicitamente do problema da estimação de proporções, P.

$$y_{ij} = I\left[(i,j) \in A\right] = \begin{cases} 1, & \text{se a unidade } j \text{ do conglorerado } i \text{ possui o atributo, } A \subset U; \\ 0, & \text{caso contrário.} \end{cases}$$

- Lembrando que **proporção é equivalente à média** de uma variável do tipo indicadora, que só pode assumir valores 0 (não possui a característica de interesse) ou 1 (possui a característica de interesse) e, novamente, é fácil derivar as expressões para estimar proporções e avaliar a precisão das estimativas a partir das expressões para estimação de média:

    - o total populacional, $T = \sum_{i \in C} T_i = N_A$,, onde N_A representa o **número de unidades populacionais** com o **atributo** de interesse;

    - a média populacional, $\overline{Y} = \frac{1}{N} \sum_{i \in C} T_i = \frac{T}{N} = \frac{N_A}{N} = P$.

Para casa

- Ler o capítulo 12 do livro 'Amostragem: Teoria e Prática Usando R'.
- Rever os slides.

Próxima aula

- Acompanhar o material no moodle.

Amostragem por Conglomerados

- Amostragem por Conglomerados em Um Estágio simples - **AC1S**
- Laboratório de R

Muito obrigado!

Fonte: imagem do livro *Combined Survey Sampling Inference: Weighing of Basu's Elephants*.

Referências

- Amostragem: Teoria e Prática Usando o R
- **Elementos de Amostragem**, Bolfarine e Bussab.
- Cochran(1977)

Resumo da notação

Abaixo segue um resumo de estimadores do total, média e respectivas variâncias sob AC1S.

Estimador	HT
Total	$\widehat{T}_{AC1S/HT} = \frac{M}{m} \sum_{i \in a} T_i = M\overline{y}_C = \sum_{i \in a} \sum_{j \in C_i} d_{ij} y_{ij}$
Média	$\overline{y}_{AC1S/HT} = \frac{\widehat{T}_{AC1S/HT}}{N} = \frac{M}{N} \frac{1}{m} \sum_{i \in a} T_i = \overline{y}_C / \overline{N}$
Variância do total	$\widehat{V}ar_{AC1S}\left(\widehat{T}_{AC1S/HT}\right) = M^2 \left(\frac{1}{m} - \frac{1}{M}\right) \widehat{S}^2_{ec}$
Variância da média	$\widehat{V}ar_{AC1S}\left(\overline{y}_{AC1S/HT}\right) = \frac{1}{\overline{N}^2} \left(\frac{1}{m} - \frac{1}{M}\right) \widehat{S}^2_{ec}$

Resumo da notação

Abaixo segue um resumo de estimadores do total, média e respectivas variâncias sob AC1S.

Estimador	Razão
Total	$\widehat{T}^{R}_{AC1S} = \frac{N}{n}\sum_{i\in a} T_i = N\overline{y} = \sum_{i\in a}\sum_{j\in C_i} w^{R}_{ij} y_{ij}$
Média	$\overline{y}^{R}_{AC1S} = \frac{\widehat{T}^{R}_{AC1S}}{N} = \frac{1}{n}\sum_{i\in a} T_i = \overline{y}_C/\overline{n} = \overline{y}$
Variância do total	$\widehat{V}ar_{AC1S}\left(\widehat{T}^{R}_{AC1S}\right) = M^2\left(\frac{1}{m}-\frac{1}{M}\right)\frac{1}{m-1}\sum_{i\in a} N_i^2(\overline{Y}_i - \overline{y})^2$
Variância da média	$\widehat{V}ar_{AC1S}\left(\overline{y}^{R}_{AC1S}\right) = \frac{1}{\overline{n}^2}\left(\frac{1}{m}-\frac{1}{M}\right)\frac{1}{m-1}\sum_{i\in a} N_i^2\left(\overline{Y}_i - \overline{y}\right)^2$