MAT02025 - Amostragem 1

AAS: amostragem aleatória simples com reposição

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

Porto Alegre, 2023

Relembrando

Relembrando

A **amostragem aleatória simples**[1] (AAS) é um processo para selecionar n unidades de N de modo que cada uma das ${}_N C_n$ amostras distintas tenha uma **chance igual de ser extraída**.

[1] Também conhecida como **amostragem casual simples** ou **amostragem acidental irrestrita**

Amostragem aleatória simples

- Como um número sorteado é removido da população em todos os sorteios subsequentes, esse método também é chamado de amostragem aleatória **sem reposição**.
- A amostragem aleatória **com reposição** é inteiramente viável: em qualquer sorteio, todos os N membros da população têm a mesma chance de serem sorteados, não importa quantas vezes eles já tenham sido sorteados.
- As fórmulas para as variâncias e variâncias estimadas das estimativas feitas a partir da amostra são frequentemente mais simples quando a amostragem é "com reposição" do que quando é "sem reposição".
    - Por esta razão, a amostragem com reposição é às vezes usada nos planos de amostragem mais complexos, embora à primeira vista pareça fazer pouco sentido em ter a mesma unidade duas ou mais vezes na amostra.

Amostragem aleatória simples com reposição

AAS com reposição

- Uma **amostra aleatória simples com reposição (AASc)** é sorteada unidade por unidade.
- As unidades da população são numeradas de 1 a N.
- Uma série de números aleatórios entre 1 e N é então sorteada, por meio de uma tabela de números aleatórios ou por meio de um **programa de computador** que produz tal tabela.
- Em qualquer sorteio, o processo usado deve dar uma chance igual de seleção a qualquer número na população.
  - Uma vez sorteada a unidade, ela é reposta na população e sorteia-se um elemento seguinte.
- Repete-se o procedimento até que n unidades tenham sido sorteadas.
  - Estas unidades **constituem a amostra (selecionada)**.

AAS com reposição

- Portanto, a probabilidade de que todas as n unidades especificadas sejam selecionadas em n sorteios é

$$\frac{1}{N}\cdot\frac{1}{N}\cdot\frac{1}{N}\ldots\frac{1}{N}=\left(\frac{1}{N}\right)^{n}=\frac{1}{N^{n}}.$$

Seleção de uma AAS com reposição

Exemplo no R

```
sample(x = 1:528, size = 10, replace = TRUE)
```

```
##  [1] 351 118 375 256 500 345 385 360 165 209
```

```
sample(x = 1:128, size = 10, replace = TRUE)
```

```
##  [1] 124 111  59 115  92  34 120  10  48  29
```

Propriedades dos estimadores na AASc

- Os estimadores $\overline{y} = \sum_{i=1}^{n} Y_i/n$ e $\widehat{Y}_T = N\overline{y}$ apresentam estimativas **não viesadas** para $\overline{Y} = \sum_{i=1}^{N} Y_i/N$ e $Y_T = \sum_{i=1}^{N} Y_i$, respectivamente.
  - **Exercício:** demonstre o resultado acima.
- As expressões das variâncias podem ser obtidas utilizando o mesmo artifício das **variáveis indicadoras** de seleção que foi utilizado no esquema de amostragem aleatório simples sem reposição.

Propriedades dos estimadores na AASc

- No caso da **AAS com reposição**, a unidade i pode aparecer $0, 1, 2, \ldots, n$ vezes na amostra.
- Seja t_i o **número de vezes que a unidade i aparece na amostra**. Então, temos

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{N} t_i Y_i.$$

Propriedades dos estimadores na AASc

- Dessa forma, t_i se distribui como uma **variável binomial**, em ***n* tentativas**, com $P = 1/N$. Portanto,

$$\mathrm{E}\left(t_i\right) = nP = \frac{n}{N}, \quad \mathrm{Var}\left(t_i\right) = nP(1-P) = n\left(\frac{1}{N}\right)\left(1-\frac{1}{N}\right).$$

- Conjuntamente, as variáveis t_i apresentam uma **distribuição multinomial** (ver, por exemplo, **Paulino et al. (2018)**[2], Apêndice). Por isso,

$$\mathrm{Cov}\left(t_i, t_j\right) = -\frac{n}{N^2}.$$

[2] Paulino, C.D., Amaral Turkman, M.A., Murteira, B., Silva, G.L. (2018). **Estatística Bayesiana**, 2ª edição. Fundação Calouste Gulbenkian, Lisboa.

Propriedades dos estimadores na AASc

- Combinando as expressões anteriores, temos, para a **amostragem aleatória simples com reposição**:

$$
\begin{aligned}
\mathrm{Var}\left(\overline{y}\right) &= \frac{1}{n^2}\left[\sum_{i=1}^{N} Y_i^2 \frac{n(N-1)}{N^2} - 2\sum_{i<j} Y_i Y_j \frac{n}{N^2}\right] \\
&= \frac{1}{nN}\sum_{i=1}^{N}(Y_i - \overline{Y})^2 = \frac{\sigma^2}{n} = \frac{N-1}{N}\frac{S^2}{n}.
\end{aligned}
$$

Propriedades dos estimadores na AASc

- A **variância** para $\widehat{Y}_T = N\overline{y}$ é dada por $\mathrm{Var}\left(\widehat{Y}_T\right) = N(N-1)\frac{S^2}{n}$.
  - **Exercício:** demonstre o resultado acima.
- **Erros padrões** são obtidos tomando-se a **raiz quadrada** destas expressões para as **variâncias**.
- **Estimativas** para os **erros padrões** podem ser obtidas utilizando a **variância amostral**, s^2, para estimar S^2.
- **Supondo normalidade** para as estimativas $\overline{y}$ e $\widehat{Y}_T$, **intervalos de confiança** podem ser construídos de forma semelhante que os intervalos construídos para AAS sem reposição **(AASs)**.

Comparação entre planos amostrais

Relembrando

Estimativa do erro padrão

As fórmulas que nos dão os **erros padrões** das estimativas da **média** ($\bar{y} = (1/n)\sum_{i=1}^{n} Y_i$) e **total** ($\hat{Y}_T = N\bar{y}$) populacionais são usadas principalmente para três finalidades:

1. **comparar a precisão** obtida por amostragem aleatória simples com a precisão dada por outros métodos de amostragem;
2. para **estimar o tamanho da amostra** necessária em um levantamento que está sendo planejado;
3. para **estimar a precisão** realmente alcançada em um levantamento que foi concluído.

Comparação entre AASc e AASs

- Quando há dois planos amostrais, é importante saber qual deles é "melhor".
  - Surge a necessidade de fixar o critério pelo qual o plano será julgado.
- Como já foi discutido anteriormente, o critério mais adotado em amostragem é o **Erro Quadrático Médio**.
  - Lembre-se que quando o estimador é não viesado, $EQM(\widehat{\theta}) = \mathrm{Var}\,(\widehat{\theta})$.
- Devido a isso, existe um conceito bastante importante, que é o chamado **efeito do planejamento (EPA, do inglês *design effect*, "deff")**, que compara a variância de um plano qualquer com relação a um plano que é considerado padrão (de referência).

Comparação entre AASc e AASs

- A estatística $\overline{y}$ é, em ambos os planos **(AASc e AASs)**, um estimador não enviesado de $\overline{Y}$. Assim,

$$EPA = \frac{\text{Var}_{AASs}(\overline{y})}{\text{Var}_{AASc}(\overline{y})} = \frac{[(N-n)/N]S^2/n}{[(N-1)/N]S^2/n} = \frac{N-n}{N-1}.$$

- Quando $EPA > 1$, tem-se que **o plano do numerador é menos eficiente que o padrão**.
- Quando $EPA < 1$, tem-se que **o plano do numerador é mais eficiente que o padrão**.

Comparação entre AASc e AASs

- Da expressão acima vê-se que

$$\frac{N-n}{N-1} \leq 1,$$

ou seja, o plano **AASs** é sempre “melhor” (mais eficiente) do que o plano **AASc**.

- Só para amostras de tamanho 1 é que os dois se equivalem.
- Note que este resultado confirma a intuição popular de que amostras sem reposição são “melhores” do que aquelas com elementos repetidos.

Para casa

- Resolver os exercícios[3] 3.1, 3.2, 3.4a, 3.4b, 3.5, 3.6, 3.7 do Capítulo 3 do livro **Elementos de Amostagem**[4] (disponível no Sabi+).
  - Trazer dúvidas para aula.

[3] PQPs.

[4] Bolfarine, H. e Bussab, W. O. **Elementos de Amostagem**, Blucher, 2005, p. 83-85.

Próxima aula

- Estimativa de um índice.

Por hoje é só!

Bons estudos!