MAT02025 - Amostragem 1

AAS: variâncias dos estimadores

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, 2021

Relembrando

Propriedades dos estimadores

Teorema 9.1

A média amostral $\overline{y}$ é uma estimativa sem tendência para $\overline{Y}$.

Variâncias dos estimadores

Variâncias dos estimadores

- Normalmente, a variância de Y_i, em uma **população finita**, é definida por

$$\sigma^2 = \frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N}.$$

Variâncias dos estimadores

- **Por uma questão de notação**, os resultados são apresentados de acordo com uma expressão ligeiramente diferente, na qual usa-se o divisor $(N-1)$ em lugar de N. Assim, temos

$$S^2 = \frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N-1}.$$

Variâncias dos estimadores

- Examinamos, agora, a variância de $\bar{y}$. Isso quer dizer $\mathrm{E}\left[(\bar{y} - \overline{Y})^2\right]$, tomado dentre todas as ${}_{N}C_{n}$.

Teorema 10.1

A variância do valor médio, $\bar{y}$, de uma amostra aleatórica simples é dada pela fórmula:

$$\mathrm{Var}\left(\bar{y}\right) = \mathrm{E}\left[(\bar{y} - \overline{Y})^2\right] = \frac{S^2}{n}\frac{(N-n)}{N} = \frac{S^2}{n}(1-f),$$

em que $f = n/N$ é a **fração de amostragem**.

Variâncias dos estimadores

Demonstração. Mais uma vez, vamos utilizar o método sugerido por **Cornfield (1944)**[1]. Neste caso, representamos a média amostral por

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{N} a_i Y_i,$$

em que o somatório abrange todas as N unidades da população.

[1] Cornfield, J. (1944). *On Samples from Finite Populations*. Journal of the American Statistical Association, 39(226), 236.

Variâncias dos estimadores

- Lembre que, nesta representação da média amostral, os a_i **são variáveis aleatórias** ($a_i = 1$, se i-ésima unidade da população é selecionada para a amostra; e $a_i = 0$, caso contrário), e os Y_i **constituem um conjunto de números fixos**.
- Ainda, temos que a_i é uma variável aleatória com **distribuição Bernolli** com

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

Variâncias dos estimadores

- A **covariância** de a_i e a_j é, por definição,

$$\mathrm{Cov}\left(a_i, a_j\right) = \mathrm{E}\left\{[a_i - \mathrm{E}\left(a_i\right)][a_j - \mathrm{E}\left(a_j\right)]\right\} = \mathrm{E}\left(a_i a_j\right) - \mathrm{E}\left(a_i\right)\mathrm{E}\left(a_j\right).$$

- Note que $a_i a_j$ é igual a 1 se as unidades i e j estiverem na amostra, ou é 0, caso contrário.
  - Logo, $\mathrm{E}\left(a_i a_j\right) = 1 \times \Pr(a_i a_j = 1) + 0 \times \Pr(a_i a_j = 0) = \Pr(a_i a_j = 1) = \Pr(a_i = 1, a_j = 1)$.

Variâncias dos estimadores

- A probabilidade de que duas unidades específicas estejam ambas na amostra é $[n(n-1)]/[N(N-1)]$. Portanto,

$$\begin{aligned}
\mathrm{Cov}\left(a_i, a_j\right) &= \mathrm{E}\left(a_i a_j\right) - \mathrm{E}\left(a_i\right)\mathrm{E}\left(a_j\right) \\
&= \frac{n(n-1)}{N(N-1)} - \left(\frac{n}{N}\right)^2 = -\frac{n}{N(N-1)}\left(1 - \frac{n}{N}\right).
\end{aligned}$$

Variâncias dos estimadores

- Agora, temos que

$$
\begin{aligned}
\mathrm{Var}\left(\overline{y}\right) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{N} a_i Y_i\right)\\
&= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{N} a_i Y_i\right)\\
&= \frac{1}{n^2}\left[\sum_{i=1}^{N}\mathrm{Var}\left(a_i Y_i\right) + 2\sum_{i<j}\mathrm{Cov}\left(a_i Y_i, a_j Y_j\right)\right]\\
&= \frac{1}{n^2}\left[\sum_{i=1}^{N} Y_i^2\mathrm{Var}\left(a_i\right) + 2\sum_{i<j} Y_i Y_j\mathrm{Cov}\left(a_i, a_j\right)\right]\\
&= \frac{1}{n^2}\left[\sum_{i=1}^{N} Y_i^2\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right) + 2\sum_{i<j} Y_i Y_j\left\{-\frac{1}{N}\frac{n}{(N-1)}\left(1-\frac{n}{N}\right)\right\}\right]
\end{aligned}
$$

Variâncias dos estimadores

$$= \frac{(1-f)}{nN}\left[\sum_{i=1}^{N} Y_i^2 - \frac{1}{N-1}\sum_{i\neq j} Y_i Y_j\right]$$

$$= \frac{(1-f)}{nN}\left[\sum_{i=1}^{N} Y_i^2 + \frac{1}{N-1}\left(\sum_{i=1}^{N} Y_i^2 - N^2\overline{Y}^2\right)\right] \text{ (soma produtos mesma variável)}$$

$$= \frac{(1-f)}{nN}\frac{N}{(N-1)}\left[\sum_{i=1}^{N} Y_i^2 - N\overline{Y}^2\right]$$

$$= \frac{(1-f)}{n(N-1)}\left[\sum_{i=1}^{N} (Y_i - \overline{Y})^2\right] \text{ (soma dos desvios quadráticos)}$$

$$= \frac{S^2}{n}(1-f).$$

- Para estas duas últimas identidades, ver o Capítulo 2.6, “Expressões úteis”, do livro “Elementos de Amostragem” **(Bolfarine e Bussab, 2005)**.

Variâncias dos estimadores

Corolário 10.1

O **erro padrão** de $\overline{y}$ é dado pela fórmula

$$\sigma_{\overline{y}} = \frac{S}{\sqrt{n}}\sqrt{(N-n)/N} = \frac{S}{\sqrt{n}}\sqrt{1-f}.$$

Variâncias dos estimadores

Corolário 10.2

A **variância** de $\hat{Y}_T = N\overline{y}$, como estimativa do valor total da população Y_T, é dada pela fórmula

$$\mathrm{Var}\,(\hat{Y}_T) = \mathrm{E}\,[(\hat{Y}_T - Y_T)^2] = \frac{N^2S^2}{n}\frac{(N-n)}{N} = \frac{N^2S^2}{n}(1-f).$$

Variâncias dos estimadores

Corolário 10.3

O **erro padrão** de $\hat{Y}_T$ é calculado pela fórmula

$$\sigma_{\hat{Y}_T} = \frac{NS}{\sqrt{n}}\sqrt{(N-n)/N} = \frac{NS}{\sqrt{n}}\sqrt{1-f}.$$

Correções das populações finitas

Correções das populações finitas

- Para uma amostra aleatória de tamanho n, proveniente de uma **população infinita**, é sabido que a variância do valor médio é σ^2/n.
- A única modificação desse valor, quando a **população é finita**, é a introdução do fator $(N-n)/N$.
- Os **fatores** $(N-n)/N$ **para a variância**, e $\sqrt{(N-n)/N}$, **para o erro padrão**, são chamados **correções das populações finitas (cpf)**.
  - São apresentados com o divisor $(N-1)$, em vez N, pelos autores que dão os resultados em função de σ.

Correções das populações finitas

- Desde que a fração n/N, permaneça baixa, esses fatores $([N-n]/N)$ se aproximam do valor da unidade, e o tamanho da população, como tal, não tem influência direta no erro padrão do valor médio amostral.
  - Ou seja, $\frac{(N-n)}{N} = 1 - \frac{n}{N} \approx 1$, quando $\frac{n}{N} \approx 0$.
  - Por exemplo, se S é o **mesmo em duas populações**, uma amostra de 500 unidades de uma população de 200.000, dá uma estimativa quase tão precisa do valor médio da população quanto uma amostra de 500 unidades de uma população de 10.000.
- Na prática, as **cpf** podem ser ignoradas, quando quer a fração de amostragem não exceda a 5% e, em muitos casos, até mesmo quando seu valor chegue a 10%.
  - **A consequência de se ignorar as correções é a superestimação do erro padrão da estimativa $\overline{y}$.**
  - **Pergunta:** no que isto implica?

Para casa

- Refaça as demonstrações da aula de hoje.
- Utilize o **Teorema 10.1** apresentado na aula para demonstrar os **Corolários 10.1**, **10.2** e **10.3**.

Próxima aula

- Estimativa do erro padrão de uma amostra e limites de confiança.

Por hoje é só!

Bons estudos!