Aula 14/08

população

Amostragem

Inferência

Amostra

Amostragem é utilizada diariamente em diversas áreas, sempre com objetivo de obter informações sobre o todo baseando-se no resultado de uma amostra

verificar se um prato esta bem temperado

Como selecionar uma "boa" amostra?

* vies de seleção

Eleições Americanas

Uma grande amostra mal selecionada é muito pior do que uma pequena amostra bem selecionada

---

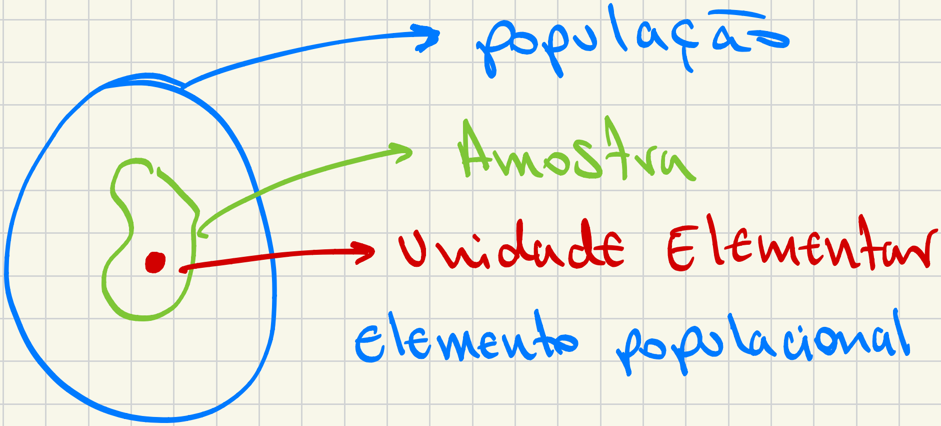
Unidade Elementar (UE): É o elemento de uma população, é o objeto portador das informações que pretendemos coletar.

Menor elemento onde podemos aplicar um tratamento observar ou medir uma certa característica

População: Conjunto de todas as unidades experimentais possíveis

Amostra: Subconjunto de unidades experimentais retiradas da população

Unidade Elementar (UE): Cada elemento individual dentro da população ou amostra.



Aula 19/03

Propósito da amostra:

Fornecer informações para descrever os parâmetros da população da melhor maneira possivel

Uma boa amostra permite a generalização de seus resultados dentro de limites aceitaveis de duvidas

Boa amostra é aquela que é representativa. Aquela que é uma microrrepresentação do universo em estudo.

Tipos de Amostragem

→ **Probabilística:** Seleção aleatória com probabilidade conhecida/permite estimar erro amostral e fazer inferências para a população

→ **Não-probabilística:** Seleciona unidades sem probabilidades conhecidas ⇒ mais sujeita a viés e não permite estimar o erro amostral de forma rigorosa

Parâmetro Populacional

é a nomenclatura utilizada para denotar o vetor correspondente a todos os valores de uma variavel de interesse que se denota por:

$$D = (y_1, y_2, \dots, y_N)$$

ou seja podemos entender um parâmetro como um número que representa uma caracteristica da população

Parâmetros Populacionais

* Total Populacional

$$\theta(D) = \theta(Y) = \tau = \sum_{i=1}^{N} y_i$$

* Média Populacional

$$\theta(D) = \theta(Y) = \mu = \bar{Y} = \frac{1}{N} \cdot \sum_{i=1}^{N} y_i$$

* Variância Populacional

$$\theta(D) = \theta(Y) = \sigma^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \mu)^2$$

$\hookrightarrow \bar{Y}$

* Covariância Populacional

$$\theta(D) = \sigma_{xy} = cov(x, y)$$

$$= \frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y) \quad (**)$$

(**) ou as vezes

$$\theta(D) = S_{xy} = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)$$

onde $\boxed{\mu_x = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i}$ média populacional

Plano Amostral

Um plano amostral é o conjunto de regras e procedimentos que especifica como será feita a seleção de uma amostra a partir de uma população

↳ como um passo organizado que define como vai ser selecionada a amostra de uma população

* Qual é a população?
* Qual é a unidade elementar?
* Tamanho da amostra?
* Método de seleção?

Amostras

$$\mathcal{U} = \{1, 2, 3, \dots, N\}$$

definição: Uma sequência de n unidades de $\mathcal{U}$ é denominada uma amostra ordenada de $\mathcal{U}$ isto é $s = (x_1, x_2, \dots, x_n)$ tal que $x_i \in \mathcal{U}$. x_i é o i-ésimo componente da amostra ordenada s.

contar o número de vezes que i aparece na amostra s

definição: Seja $f_i(s)$ a variável que indica o número de vezes (frequência) que a i-ésima unidade populacional aparece na amostra s.

definição: Seja $\delta_i(s)$ a variável binária que indica a presença da i-ésima unidade em s.

$$\begin{cases} \delta_i(s) = 1 & \text{se } i \in s \\ \delta_i(s) = 0 & \text{se } i \notin s \end{cases}$$

↪ Variável Indicadora

definição: A soma das frequências das unidades populacionais na amostra é denominada tamanho amostral

$$n(s) = \sum_{i=1}^{N} f_i(s)$$

soma percorre toda população mas $f_i(s)$ filtra

definição: O número de unidades populacionais distintas na amostra é denominado tamanho efetivo da amostra

$$\nu(s) = \sum_{i=1}^{N} \delta_i(s)$$

Aula 21/08

Estimadores e suas propriedades

Objetivo da Amostragem: Estimar o parâmetro populacional desconhecido $\theta(d)$ utilizando uma estatística com boas propriedades

Um estimador é uma expressão estatística que irá estimar o parâmetro populacional enquanto que uma estimativa é o valor numérico do estimador para uma dada amostra

Um estimador $\hat{\theta}(d_s)$ é dito não-viciado segundo um plano amostral A se

$$E_A(\hat{\theta}) = \theta$$

Definição: O viés do estimador $\hat{\theta}(ds)$ segundo um plano amostral A é dado por

$$B_A(\hat{\theta}) = \mathbb{E}[\hat{\theta} - \theta] = \mathbb{E}(\hat{\theta}) - \theta$$

Definição: O Erro Quadrático Médio (EQM) do estimador $\hat{\theta}(ds)$ segundo um plano amostral A é dado por

$$EQM_A[\hat{\theta}] = \mathbb{E}[(\hat{\theta}-\theta)^2]$$

$$= \mathbb{E}[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2]$$

$$= \mathbb{E}[\hat{\theta}^2] - \mathbb{E}[2\hat{\theta}\theta] + \mathbb{E}(\theta^2)$$

$$= \mathbb{E}(\hat{\theta}^2) - 2\theta \cdot \mathbb{E}(\hat{\theta}) + \theta^2 \checkmark$$

esperança é linear e θ é um número $E(\theta) = \theta$

Somar zero

$$= \mathbb{E}(\hat{\theta}^2) - \{\mathbb{E}(\hat{\theta})^2 - \mathbb{E}(\hat{\theta})^2\} - 2\theta\,\mathbb{E}(\hat{\theta}) + \theta^2$$

$$= \underbrace{\mathbb{E}(\hat{\theta}^2) - [\mathbb{E}(\hat{\theta})]^2}_{var(\hat{\theta})} - \underbrace{[\mathbb{E}(\hat{\theta})]^2 - 2\theta\,\mathbb{E}(\hat{\theta}) + \theta^2}_{(\mathbb{E}(\hat{\theta}) - \theta)^2}$$

$$var(\hat{\theta}) + \{B(\hat{\theta})\}^2$$

Refazer

Erro Quadrático Médio

$\theta \in \mathbb{R}$ é um número

$$EQM(\hat\theta) = E((\hat\theta - \theta)^2)$$
$$= E(\hat\theta^2 - 2\hat\theta\theta + \theta^2)$$
$$= E(\hat\theta^2) - E(2\hat\theta\theta) + \theta^2$$
$$= E[\hat\theta^2] - 2\theta E(\hat\theta) + \theta^2 - [E(\theta)^2]$$

$$= \underbrace{E(\hat\theta^2) - [E(\theta)^2]}_{var(\theta)} + E[(\theta)^2]$$
$$\underbrace{- 2\theta E(\hat\theta) + \theta^2 + E[(\hat\theta)^2]}$$

$$var(\hat\theta) + E[(\hat\theta - \theta)^2]$$

$$var(\hat\theta) + (B(\hat\theta)^2) \longrightarrow (\text{vício}/\text{viés})^2$$

$$var(\hat\theta) = E[\hat\theta^2] - [E(\theta)^2]$$

Refazer novamente

Acho pouco provável cair na prova mas é uma demonstração técnica que ajuda vício para um estimador θ

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

$$(B(\hat{\theta}))^2 = (E(\hat{\theta}) - \theta)^2$$
$$= [E(\hat{\theta})^2] - 2\theta E(\hat{\theta}) + \theta^2$$

E

$$EQM(\hat{\theta}) = E((\hat{\theta} - \theta)^2)$$ = variancia do estimador + um (viés/vício)^2

$$= E(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2)$$

esperança é linear e $\boxed{E(\theta) = \theta}$???

$$= E(\hat{\theta}^2) - E(2\hat{\theta}\theta) + E(\theta^2)$$ ✓

$$= E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2$$

$var(\hat{\theta}) = E(\hat{\theta}^2) - [E(\hat{\theta})^2]$

$$= \boxed{E(\hat{\theta}^2)} - 2\theta E(\hat{\theta}) + \theta^2 + [E(\hat{\theta})^2] - \boxed{[E(\hat{\theta})^2]}$$ somam zero

$var(\hat{\theta})$

$$= var(\theta) \underbrace{- 2\theta E(\hat{\theta}) + \theta^2 + [E(\hat{\theta})^2]}_{\text{vies/vicio } [B(\hat{\theta})]^2}$$

ou seja o Erro Quadrático Médio

$$EQM[\hat{\theta}] = Var(\hat{\theta}) + (B(\hat{\theta}))^2$$

→ vies

Resumo:

Estimador: É uma função construida a partir da amostra que busca estimar o parâmetro

Ou seja uma função que aplicamos aos dados amostrais $(X_1, X_2, \ldots, X_n)$ para produzir uma aproximação (uma estimativa) de um parâmetro populacional

Estimativa: É o valor númerico especifico que resulta da aplicação do estimador a uma determinada amostra

Estimador não-viciado

$$E(\hat{\theta}) = \theta$$

// possui vício igual a zero

Um bom estimador deve ser

* não-viciado: Não distorce o parâmetro

* consistente: Quanto maior a amostra mais próximo do parâmetro real

* Eficiente: Possui a menor variância possível entre os estimadores não-viciados

Viés/Vício

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

$$EQM(\hat{\theta}) = E((\hat{\theta}-\theta)^2)$$

$$= var(\hat{\theta}) + (B(\hat{\theta}))^2$$

Erro Quadrático Médio

Combina ambos os conceitos (vies e variância) em um único número.
Mede a qualidade geral do estimador ou seja o quão "perto" em média nossas estimativas estão do real valor

$$EQM(\hat{\theta}) = Var(\hat{\theta}) + (B(\hat{\theta})^2)$$

→ vies

O que é $f_i(s)$?

A variável $f_i(s)$ é a frequência de inclusão da i-ésima unidade elementar na amostra s.

Em outras palavras

$f_i(s)$ = número de vezes que a unidade i aparece em s

A distribuição de $f_i(s)$ depende do plano amostral

Amostra Aleatória Simples Sem Reposição $(AAS)_s$:

Cada unidade pode aparecer somente uma vez na amostra

Portanto $f_i(s)$ é uma variável aleatória Bernoulli

$$f_i(s) = \begin{cases} 1 & \text{se } i \text{ esta em } s \\ 0 & \text{caso contrario} \end{cases}$$

A probabilidade de inclusão é

$$\pi_i = P(f_i(s) = 1) = \frac{n}{N}$$

Amostra Aleatória Simples Com Reposição (AAS)c

Uma mesma unidade pode aparecer mais de uma vez na amostra

Portanto $f_i(s)$ nesse caso é uma variável aleatória com distribuição Binomial:

$$f_i(s) \sim Binomial(n, p_i)$$

onde p_i é a probabilidade de selecionar a unidade i em cada sorteio (geralmente $p_i = \frac{1}{N}$ para (AAS)c

Podemos reescrever a média amostral $s = (y_1, y_2, \dots)$ → amostra

A média amostral usual é

$$\bar{y} = \frac{1}{n} \cdot \sum_{i \in s} y_i$$

Porem a soma $\sum_{i \in s} y_i$ pode ser reescrita como uma soma sobre toda a população ponderada pela sua frequência de inclusão

$$\sum_{i \in s} y_i = \sum_{i=1}^{N} y_i \, f_i(s)$$

filtra naturalmente as unidades que estão em s

Se a unidade i não está na amostra $f_i(s) = 0$ não contribui com a soma (**)

muito muito Importante

(**) Assim podemos reescrever a média amostral como:

$$\bar{y}_s = \frac{1}{n} \cdot \sum_{i=1}^{N} y_i \cdot f_i(s)$$ filtra na amostra

---

Exemplo

$U = \{1, 2, 3, 4, 5, 6\}$

$Y = \{10, 20, 20, 5, 5, 4\}$

Considere a amostra $s = (3, 4, 6, 6)$
$(n = 4)$

$$\bar{y}_s = \frac{1}{4} \cdot \sum_{i \in s} y_i = \frac{1}{4}\left(y_3 + y_4 + y_6 + y_6\right)$$

$$= \frac{1}{4}\left(20 + 5 + 4 + 4\right) = \frac{33}{4} = 8{,}25$$

$\Rightarrow$

⇒ ou ainda

$$h(d_s) = \bar{y}_s = \frac{1}{n} \cdot \sum_{i=1}^{N} y_i f_i(s)$$

Aula 26/08

Amostragem Aleatória Simples (AAS)

Definição: A (AAS) é um plano amostral em que selecionam n unidades de uma população de tamanho N de modo que todas as combinações possíveis de n unidades tenham a mesma probabilidade de serem escolhidas

$(AAS)_c$: Amostragem Aleatoria Simples com reposição

$(AAS)_s$: Amostragem Aleatória Simples sem reposição

Observações:

* Em geral a $(AAS)_s$ é a mais natural, pois não faz sentido selecionar o mesmo elemento mais de uma vez

* No entanto a $(AAS)_c$ é muito utilizada em demonstrações teóricas pois facilita o estudo das propriedades dos estimadores.

Ao selecionar n elementos de uma população de tamanho N o número de amostras possíveis depende do tipo de AAS e se a ordem importa ou não

AAS com reposição: Cada uma das n posições pode ser ocupada por qualquer um dos N elementos

N^n possíveis amostras

AAS sem reposição:

→ caso não ordenado ou seja apenas importa quais elementos foram escolhidos

$\binom{N}{n}$ possíveis amostras $\quad P(s) = \frac{1}{\binom{N}{n}}$

→ caso ordenado: Agora importa a ordem dos elementos escolhidos

$$A_{N,n} = \frac{N!}{(N-n)!}$$

Anagramas com letras iguais

$P(s) = \frac{1}{A_{N,n}}$

Distribuição da v.a f_i (AAS)$_c$

f_i: Indica o número de vezes que o elemento i aparece na amostra de tamanho n

A probabilidade do elemento i ser selecionado é $1/N$ então

$$f_i \sim \text{Binomial}(n, 1/N)$$

$$E[f_i] = n \cdot \frac{1}{N} = \frac{n}{N}, \quad \forall\, i = 1, 2, \dots, N$$

$q = 1 - p$

$$\text{var}[f_i] = n \cdot p\, q = n \cdot \frac{1}{N} \cdot \left(1 - \frac{1}{N}\right)$$

$$= \frac{n}{N}\left(1 - \frac{1}{N}\right) \quad \forall\, i = 1, 2, \dots, N$$

Parâmetros de Interesse

Total Populacional: $\tau = \sum_{i=1}^{N} y_i$

Média Populacional: $\mu = \frac{1}{N} \sum_{i=1}^{N} y_i$

Variância Populacional:

* $\sigma^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \mu)^2$

ou $S^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (y_i - \mu)^2$

Teorema: A estatística $t(s)$ total da amostra é definida por

$$t(s) = \sum_{i \in s} y_i$$

total da amostra

muito muito útil também

$t(s) = \sum_{i \in s} y_i$ tem para o plano (AAS)c as seguintes propriedades

$$E(t) = n \cdot \mu$$
$$Var(t) = n \cdot \sigma^2$$

$E(f_i(s)) = \frac{n}{N}$

Entender melhor (I)

Prova:

$$t(s) = \sum_{i \in s} y_i = \sum_{i=1}^{N} y_i \cdot f_i(s)$$

é um número

$$E(t) = E\left[\underbrace{\sum_{i=1}^{N} y_i}_{\text{total populacional}} \cdot f_i(s)\right] = E[f_i(s)] \cdot \sum_{i=1}^{N} y_i$$

$$= \frac{n}{N} \cdot \sum_{i=1}^{N} y_i$$

média populacional

$$= n \cdot \underbrace{\frac{1}{N} \cdot \sum_{i=1}^{N} y_i}_{\mu} = n \cdot \mu$$

$$\Rightarrow E(t) = n \cdot \mu$$

$$t(s) = \sum_{i \in s} y_i \quad \text{total da amostra}$$

No caso de uma amostra $(AAS)_c$ temos

$$E(t) = n \cdot \mu \qquad \begin{cases} E(f_i(s)) = \dfrac{n}{N} \\[2ex] var(f_i(s)) = \dfrac{n}{N}\left(1 - \dfrac{1}{N}\right) \end{cases}$$

$$var(t) = n\sigma^2$$

$$E(t) = E\left(\sum_{i \in s} y_i\right) = E\left(\sum_{i=1}^{N} y_i \cdot f_i(s)\right)$$

$$= E(f_i(s)) \cdot \sum_{i=1}^{N} y_i$$

$$= \underbrace{\frac{n}{N} \cdot \sum_{i=1}^{N} y_i}$$

μ média populacional

$$= n\mu$$

Made with Goodnotes

Vamos agora tentar encontrar a var(t(s))

$$\boxed{t(s) = \sum_{i \in s} y_i}$$

total da amostra

ou ainda

$$\boxed{t(s) = \sum_{i=1}^{N} y_i \cdot f_i(s)}$$

somamos sobre toda a população

e ponderamos pela frequencia em que aparecem na amostra $f_i(s)$

$$\bar{y} = \frac{1}{n} \cdot \sum_{i \in s} y_i$$

media amostral

$$\bar{y} \cdot n = \sum_{i \in s} y_i = t(s)$$

ou seja

$$\bar{y} = \frac{t(s)}{n}$$

Agora queremos encontrar a variancia de $t(s)$

$$t(s) = n \cdot \bar{y} \qquad \bar{y} = \frac{1}{n} \cdot \sum_{i \in s} y_i$$

$$var(t) = Var(n \cdot \bar{y}) \qquad \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{N} y_i \cdot f_i(s)$$

$$= n^2 \cdot var(\bar{y}) \quad \frac{\sigma^2}{n}$$

$$= n^{\cancel{2}} \cdot \frac{\sigma^2}{\cancel{n}} = n\,\sigma^2$$

Falta organizar um pouco a ordem de alguns tópicos mas o fato é que

$$var(\bar{y}) = \frac{\sigma^2}{n}$$

$$\bar{y} = \frac{1}{n} \cdot \sum_{i \in S} y_i$$ média amostral

$$E(\bar{y}) = E\left(\frac{1}{n} \cdot \sum_{i \in S} y_i\right)$$

$$= E\left(\frac{1}{n} \cdot \sum_{i=1}^{N} y_i \cdot f_i(s)\right)$$

$$= \frac{1}{n} \cdot \sum_{i=1}^{N} y_i \cdot E(f_i(s))$$

no caso de uma $(AAS)_c$

$$E(f_i(s)) = \frac{n}{N}$$

$$= \cancel{\frac{1}{n}} \cdot \frac{\cancel{n}}{N} \cdot \sum_{i=1}^{N} y_i = \frac{1}{N} \cdot \sum_{i=1}^{N} y_i = \mu$$

$E(\bar{y}) = \mu \Rightarrow$ media populacional

$$var(\bar{y}) = var\left(\frac{1}{n} \cdot \sum_{i \in S} y_i\right)$$

$$= \frac{1}{n^2} \cdot var\left(\sum_{i=1}^{N} y_i \cdot f_i(s)\right)$$

$$\frac{n \cdot \sigma^2}{n^2}$$

$$= \frac{1}{n^2} \cdot \mu^2 \cdot var(f_i(s))$$

$$= \frac{\sigma^2}{n}$$

$$\circledast \quad \frac{1}{n^2} \cdot \tau^2 \cdot \operatorname{var}(f_i(s))$$

no caso de uma (AAS)c

$\operatorname{var}(f_i(s)) =$

$\frac{n}{N}\left(1 - \frac{1}{N}\right)$

$$\frac{\tau^2}{n^{\cancel{2}}} \cdot \cancel{n} \cdot \frac{1}{N}\left(1 - \frac{1}{N}\right)$$

$$\frac{\tau^2}{n} \cdot \frac{1}{N}\left(1 - \frac{1}{N}\right)$$

$$\frac{\tau}{N} = \mu$$

$$= \frac{\tau^2}{n \cdot N} - \overbrace{\frac{\tau^2}{N^2}}^{\mu^2} \cdot \frac{1}{n} = \overbrace{\frac{\tau}{N}}^{\mu} \cdot \frac{\tau}{n} - \frac{1}{n} \cdot \mu^2$$

$E(\bar{y}) = \mu$

$$= \mu \cdot \frac{\tau}{n} - \frac{1}{n} \cdot \mu^2$$

$$\frac{\mu}{n}(\tau - \mu)$$

Estimação do total e da média populacionais

Vamos obter os estimadores não-viesados para τ e μ

Teorema:

A média amostral

$$\bar{y} = \frac{1}{n} \cdot \sum_{i \in s} y_i = \frac{t(s)}{n} = \hat{\mu}$$

é um estimador não-viesado da média populacional μ considerando um plano amostral $(AAS)_c$ e

$$Var(\bar{y}) = \frac{\sigma^2}{n} \quad ; \quad E(\bar{y}) = \mu$$

Prova:

Entender (II)

Queremos provar que $E(\bar{y}) = \mu$

ou seja $\bar{y}$ é não-viesado

$$E(\bar{y}) = E\left(\frac{1}{n} \cdot \sum_{i \in s} y_i\right) = E\left(\frac{t(s)}{n}\right)$$

$$= \frac{1}{n} \cdot E(\underbrace{t(s)}_{n \cdot \mu}) = \mu \quad \blacksquare$$

$t(s)$

$$= \frac{1}{n} \cdot E\left(\sum_{i \in s} y_i\right) = \frac{1}{n} \cdot E\left(\underbrace{\sum_{i=1}^{N} y_i}_{\text{total populacional}} \cdot f_i(s)\right)$$

$$= \frac{1}{n} \cdot E(\underbrace{f_i(s)}_{\frac{n}{N}}) \cdot \tau$$

$$= \cancel{\frac{n}{N}} \cdot \cancel{\frac{1}{n}} \cdot \sum_{i=1}^{N} y_i = \underbrace{\frac{1}{N} \cdot \sum_{i=1}^{N} y_i}_{\mu}$$

Refazer com atenção

$$\text{var}(\bar{y}) = \text{var}\left(\frac{1}{n} \cdot \sum_{i \in s} y_i\right)$$

$t(s)$

$$= \frac{1}{n^2} \cdot \text{var}\left(\sum_{i \in s} y_i\right)$$

total da amostra

$n\sigma^2$

$$= \frac{1}{n^2} \cdot \text{var}(t(s)) = \frac{1}{\cancel{n^2}} \cdot \cancel{n} \cdot \sigma^2$$

$$= \frac{\sigma^2}{n}$$

Refazer

Justamente o que eu não consegui provar

No plano (AASc) um estimador não viesado para o total populacional τ é dado por

$$\hat{\tau} = N\bar{y} = \frac{N}{n} \cdot t(s)$$

com variância amostral dada por

$$\text{var}(\hat{\tau}) = \frac{N^2 \cdot \sigma^2}{n}$$

Teorema: No plano (AAS)c um estimador não viesado da variância populacional σ^2 é dado por:

$$S^2 = \frac{1}{n-1} \cdot \sum_{i \in s} (y_i - \bar{y})^2$$

Prova:

$$(n-1)\, S^2 = \sum_{i \in s} (y_i - \bar{y})^2$$

não consegui entender ...

SLIDE Aula 26/08 final

$$t(s) = \sum_{i \in s} y_i$$ Total da Amostra

tem para o plano (AAS)c as seguintes propriedades

$$E(t(s)) = n \cdot \mu \qquad E(f_i(s)) = \frac{n}{N}$$

$$E(t(s)) = E\left(\sum_{i \in s} y_i\right)$$

$$= E\left(\sum_{i=1}^{N} y_i \cdot f_i(s)\right)$$

$$= E(f_i(s)) \cdot \sum_{i=1}^{N} y_i$$

$$\frac{n}{N} \cdot \sum_{i=1}^{N} y_i = n \cdot \frac{1}{N} \cdot \sum_{i=1}^{N} y_i$$

média populacional μ

$$= n \cdot \mu$$

$$\therefore \boxed{E(t) = n \cdot \mu}$$

fiz várias vezes para entender

Made with Goodnotes

$$Var(t) = n \cdot \sigma^2$$

$$E(t) = n \cdot \mu$$

$$Var[t(s)] = Var(t) \qquad t(s) = \sum_{i \in s} y_i$$

Vamos reescrever assim

$$Var(t) = E(t^2) - [E(t)]^2$$

Slides 21

não entendi

$$= E\left[\left(\sum_{i \in s} y_i\right)^2\right] - [n \cdot \mu]^2$$

???

$$= E\left[\left(\sum_{i=1}^{N} y_i \cdot f_i(s)\right)^2\right] - n^2 \cdot \mu^2$$

$$\left[\sum_{i=1}^{N} y_i \cdot f_i(s)\right]^2 = \left[y_1 \cdot f_1(s) + \dots + y_N f_N(s)\right]^2$$

Não consegui desenvolver

tem uma demonstração mais simples

Made with Goodnotes

Relembrando

Estimadores de μ e τ

$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{N} y_i \cdot f_i(s)$$

* Média Amostral

$$\bar{y} = \frac{1}{n} \cdot \sum_{i \in s} y_i$$

é um estimador não viesado da média populacional μ considerando o plano amostral (AAS)c

e $var(\bar{y}) = \frac{\sigma^2}{n}$

$$E(\bar{y}) = E\left(\frac{1}{n} \cdot \sum_{i \in s} y_i\right) = \frac{1}{n} \cdot E\left(\sum_{i=1}^{N} y_i \cdot f_i(s)\right)$$

$$= \cancel{\frac{1}{n}} \cdot E(f_i(s)) \cdot \sum_{i=1}^{N} y_i$$

$$\Rightarrow \cancel{\frac{n}{N}}$$

$$\frac{1}{N} \cdot \sum_{i=1}^{N} y_i \rightarrow \mu$$

média populacional

$$E(\bar{y}) = \mu$$

Made with Goodnotes

$$\operatorname{var}(\bar{y}) = \operatorname{var}\left(\frac{1}{n} \cdot \sum_{i \in s} y_i\right)$$

Provar isso

$$= \frac{1}{n^2} \cdot \operatorname{var}\left(\sum_{i=1}^{N} y_i \cdot f_i(s)\right)$$

$$= \frac{1}{n^2} \cdot \operatorname{var}(f_i(s)) \cdot \left(\sum_{i=1}^{N} y_i\right)^2$$

???

$$\frac{n}{N} \cdot \left(1 - \frac{1}{N}\right)$$

$$= \frac{1}{n^{\cancel{2}}} \cdot \frac{\cancel{n}}{N} \left(1 - \frac{1}{N}\right)\left(\sum_{i=1}^{N} y_i\right)^2$$

???

$$= \frac{1}{n \cdot N}\left(1 - \frac{1}{N}\right)\left(\sum_{i=1}^{N} y_i\right)^2$$

$$= \left(\frac{1}{n \cdot N} - \frac{1}{n \cdot N^2}\right) \sum_{i=1}^{N} y_i \quad \mu^2$$

$$= \frac{\mu}{n \cdot N} - \frac{\mu}{n N^2} = \frac{\mu}{n} - \frac{\mu}{n \cdot N}$$

$$= \frac{\mu}{n}\left(1 - \frac{1}{N}\right) \qquad \operatorname{var}(\bar{y}) = \frac{\sigma^2}{n}$$

Aula 28/08

Teorema: A média amostral

$\bar{y} = \frac{1}{n} \cdot \sum_{i \in s} y_i$ é um estimador

não viesado da média populacional μ considerando o plano amostral (AAS)c e $Var(\bar{y}) = \frac{\sigma^2}{n}$

Corolário: No plano (AAS)c um estimador não-viesado para o total populacional τ é dado por $T(s) = \hat{\tau} = N\bar{y}$

com $Var(T) = N^2 \frac{\sigma^2}{n}$

Importante

$$\bar{y} = \frac{1}{n} \cdot \sum_{i \in s} y_i \quad \text{I}$$

$$\bar{y} = \frac{1}{n} \cdot t(s) \quad \text{II}$$

onde $t(s) = \sum_{i \in s} y_i$

total da amostra

$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{N} y_i \cdot f_i(s) \quad \text{III}$$

da equação II temos que $t(s)$ pode ser reescrito como

$$t(s) = n \cdot \bar{y}$$

Ou seja podemos escrever $\bar{y}$ ou $t(s)$ da maneira que for mais conveniente a depender do contexto. Ajuda muito para provar/calcular ...

$var(\bar{y}) = ?$ $\quad$ $var(t(s)) = ?$

Teorema do limite Central - (AAS)c

O teorema do limite central (TLC) para n suficientemente grande considerando o plano amostral (AAS)c temos

$$\frac{\bar{y} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \overset{a}{\sim} N(0,1)$$

e

$$\frac{T - \tau}{N\sqrt{\frac{\sigma^2}{n}}} \overset{a}{\sim} N(0,1)$$

Utilizando a normalidade assintotica podemos escrever

$$P\left(\frac{|\bar{y} - \mu|}{\sqrt{\frac{\sigma^2}{n}}} \leq z_{\alpha/2}\right) \approx 1-\alpha$$

Intervalo de confiança para μ

$$P\left(\bar{y} - z_{\alpha/2}\sqrt{\frac{s^2}{n}} \leq \mu \leq \bar{y} + z_{\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}\right) \approx 1-\alpha$$

daí obtemos um intervalo de $100(1-\alpha)\%$ de confiança para μ

$$IC[\mu; 100(1-\alpha)\%]$$
$$= \left(\bar{y} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s^2}{n}} \; ; \; \bar{y} + z_{\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}\right)$$

Em palavras para construir um intervalo de confiança para a média populacional μ em amostragem com uma (AAS)c precisamos de duas estimativas

$\bar{y}$ média amostral $\Rightarrow \bar{y} = \frac{1}{n} \cdot \sum_{i \in s} y_i$

s^2 variância amostral $\Rightarrow s^2 = \frac{1}{n-1} \cdot \sum_{i \in s} (y_i - \bar{y})^2$

$$IC(\mu) = \bar{y} \pm z_{\frac{\alpha}{2}} \cdot EP(\bar{y})$$

$$EP(\bar{y}) = \sqrt{\frac{s^2}{n}}$$ erro padrão

$$s^2 = \frac{1}{n-1} \cdot \sum_{i \in s} (y_i - \bar{y})^2$$

usamos s^2 para estimar σ^2

---

*Estudar sobre intervalos de confiança

* tamanho da amostra?
como encontrar
o n?

aula 28/08

Exemplo:

$U = \{1, 2, 3\}$

$d = (12; 30; 18)$

Um plano (AAS)c com $n = 10$ foi adotado e obtivemos os seguintes valores da venda

$$d_s = (12, 18, 12, 18, 18, 30, 12, 12, 18, 30)$$

- Encontre uma estimativa da venda bruta familiar média na população e uma estimativa para a variância da estimativa obtida

* Encontre um $IC(\mu; 100(100-5)\%)$

Sabemos que um estimador não viesado para a renda bruta familiar média na população é $\bar{y}$

$$\bar{y} = \frac{1}{n} \cdot \sum_{i \in s} y_i = 18 \therefore \hat{\bar{y}} = 18$$

$$s^2 = \frac{1}{n-1} \cdot \sum_{i \in s} (y_i - \bar{y})^2$$

feito no R

$$= \frac{1}{9} \cdot \sum_{i \in s} (y_i - 18)^2 = \dots = 48$$

$$\widehat{var}(\bar{y}) = 4{,}8$$

Aula 02/09

Amostragem Aleatoria Simples sem reposição (AAS)s

* A população é numerada de 1 a N ; $\mathcal{U} = \{1, 2, \ldots, N\}$

* Sorteia-se com igual probabilidade uma das N unidades da população

sem reposição

↳ Sorteia-se o seguinte sem devolver o anterior de volta a população

* Repete-se o procedimento até que n unidades tenham sido sorteadas

Distribuição da v.a f_i

* Número de vezes que o elemento i aparece na amostra

* Sem reposição f_i assume valores 0 ou 1

* O número de amostras possíveis de tamanho n de uma população $U = \{1, 2, \ldots, N\}$

$$\binom{N}{n}$$

* Cada amostra tem probabilidade $1/\binom{N}{n}$ de ocorrer

* Qual a probabilidade do elemento i estar na amostra?

$$\pi_i = n/N$$

$$\pi_i = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

Ao final de n seleções o número de vezes que o elemento i foi sorteado (f_i) segue uma Bernoulli $\left(\frac{n}{N}\right)$

$$f_i \sim \text{Bernoulli}\left(\frac{n}{N}\right)$$

$$P(f_i = 1) = \frac{n}{N} \quad \text{e} \quad P(f_i = 0) = 1 - \frac{n}{N}$$

$$\boxed{E[f_i] = \frac{n}{N} \quad var(f_i) = \frac{n}{N}\left(1 - \frac{1}{N}\right)}$$

Estatistica $t(s)$

* Para o plano amostral
→ (AAS)c tem se:

$E(t) = n \cdot \mu$
$var(t) = n \cdot \sigma^2$

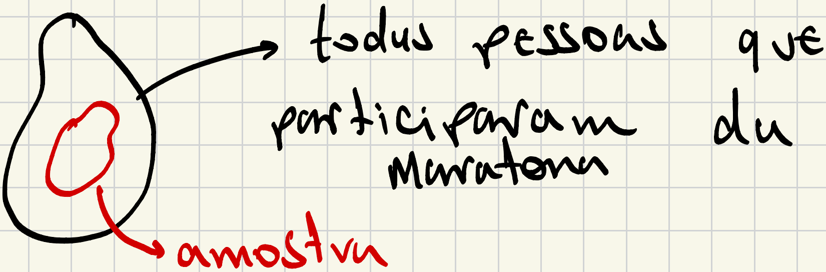
* Para o plano amostral
→ (AAS)s tem se:

$E(t) = n \cdot \mu$
$var(t) = n(1-f)s^2$ onde
$f = \frac{n}{N}$ é chamado fração amostral

e $(1-f)$ é chamado fator de correção para população finita

lista I

① Estimar o tempo médio de conclusão da Maratona de Nova York



a) Unidade de Pesquisa

Podemos entender cada atleta/corredor individual como a unidade de pesquisa.

b) População

Nesse contexto minha população seriam todas as pessoas que participaram dessa Maratona (concluíram a prova)

e) Instrumento a coleta de dados

Podemos obter os dados dos participantes utilizando um sistema de cronometragem eletrônica (chips, tags..)

2 Apresentando uma questão de pesquisa ligada a minha área que pode ser respondida com um levantamento amostral.

Queremos estimar a renda média dos clientes do Itaú Personallité.

Vamos desenhar o plano amostral

a) Unidade Elementar: É o titular da conta do Itaú Personallité

b) A população é composta por todos os titulares principais do Itaú Personnalité

③ Etapas de um levantamento amostral

(I) Definição do problema de pesquisa

→ Explicitar o que se quer medir

→ ter um objetivo claro

(II) População - Alvo

→ Definir de forma clara quem faz parte e quem não faz parte

(III) $\underbrace{\text{Unidade de Pesquisa}}$ e Resposta

elemento sobre o qual coletamos dados

IV Plano amostral

→ Escolher um método de seleção

AAS - Amostra Aleatoria Simples

Estratificada - Divide em grupos

conglomerados - Seleciona grupos

Justificar a escolha conforme o problema

V Tamanho da amostra

→ Definir n para garantir uma precisão estatística

VI Instrumento de coleta

Lista II

a) $\mu_y = \frac{1}{N} \overbrace{\sum_{i=1}^{N} y_i}^{\tau} = \frac{1}{10} \cdot \sum_{i=1}^{10} y_i$

$$= \frac{1}{10}\left(\underbrace{2+1+3+1+4+2+2+1+3+1}_{20}\right)$$

$$= 2$$

$\tau_y = 20$

$$S_y^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (y_i - \mu_y)^2$$

$$= \frac{1}{9} \cdot \sum_{i=1}^{10} (y_i - 2) = \frac{1}{9}\left(1+1+1+1+4+1+1\right)$$

$$= \frac{10}{9}$$

Tem que calcular

μ_x τ_x e S_x^2 depois

faço

② $U = \{1, 2, 3\}$

$S(U) \Rightarrow$ Espaço amostral com amostras de tamanho 2 e sem reposição

a) Queremos as distribuições das variaveis f_i e δ_i

$\underline{3} \times \underline{2} = 6$

$\# S(U) = 6$

f_i é a frequência

δ_i Indicação se esta na amostra ou não

Plano 1

$P(12) = P(13) = P(21) = P(23) = \ldots = \frac{1}{6}$

Todas amostras possuem a mesma chance de ser escolhida

f_i é uma Bernoulli$\left(\frac{2}{3}\right)$ pois

$f_1(s) = \frac{4}{6} = \frac{2}{3}$ $\quad E(f_i(s)) = \frac{2}{3}$

$f_2(s) = \frac{4}{6} = \frac{2}{3}$ $\quad f_i \sim \text{Bernoulli}\left(\frac{2}{3}\right)$

$f_3(s) = \frac{2}{3}$ $\quad \delta_i \sim \text{Bernoulli}\left(\frac{2}{3}\right)$

$f_i = 1$ é $\frac{2}{3}$ para os 3 números

$E(f_i(s)) = \frac{2}{3}$

$var(f_i(s)) = 1 - \frac{2}{3}$

③ $U = \{1, 2, 3, 4, 5\}$ $(AAS)_s$

$D = (2, 5, 7, 8, 11)$

$$\# S(U) = \frac{5 \times 4}{2} = 10$$

pois a ordem não importa

$$P(12) = P(21) = \ldots P(15) = \frac{1}{10}$$

$f_i \sim \text{Bernoulli}\left(\frac{2}{5}\right)$

$$\text{var}(f_i) = \frac{2}{5}\left(1 - \frac{2}{5}\right)$$

$$= \frac{2}{5} - \frac{4}{25} = \frac{50 - 20}{125} = \frac{30}{125} = \frac{6}{25}$$

$$t(s) = \sum_{i \in s} y_i = \sum_{i=1}^{N} y_i \cdot f_i(s)$$

② $U = \{1, 2, 3\}$

Alguns exercícios do livro

3.1

População $N=6$ tem se $U=\{1,2,\dots,6\}$

$D=(8,2,2,11,4,7)$

Um plano $(AAS)_s$ de tamanho

$n=2$ é adotado

a. Encontre a distribuição

de $\bar{y}$ e Mostre que $E(\bar{y})=\mu$

① Vamos calcular a média populacional

$$\mu=\frac{1}{N}\cdot\sum_{i=1}^{N} y_i = \frac{17}{3} \approx 5{,}66$$

(I) listando todas as amostras possíveis de tamanho $n=2$

(*) Vamos assumir que a ordem não importa eu seja

$$\binom{6}{2} = \frac{6!}{2!(6-2)!} = \frac{6 \cdot 5 \cdot \cancel{4!}}{2! \cdot \cancel{4!}} = 15$$

Número de amostras possíveis

O número total de amostras de tamanho n que podemos extrair de uma população de tamanho N (SEM REPOSIÇÃO e SEM CONSIDERAR A ORDEM)

é dado pelo coeficiente binomial

$$\text{Número de Amostras} = \binom{N}{n} = \frac{N!}{n!\,(N-n)!}$$

09/09 - Segunda

$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1} y_i = \sum_{i=1}^{N} y_i f_i$$

entender como chegar nessas contas

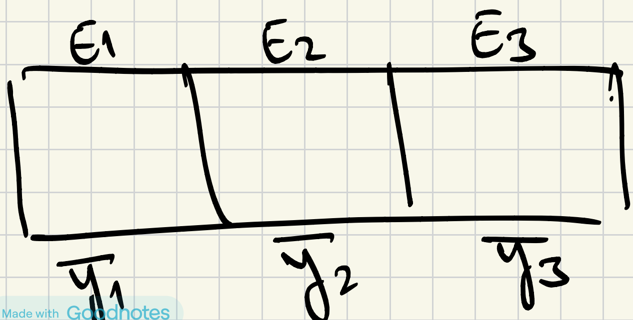
$$E(\bar{y}) = \sum_{i=1}^{N} \frac{y_i E(f_i)}{n}$$

$var(\bar{y})$ não vai ser pedido na prova

Amostragem Estratificada

Dividir a população em grupos denominados estratos

O mesmo plano amostral em cada estrato.



$$\bar{y} = \sum_{i=1}^{3} P_i \bar{y}_i = P_1\bar{y}_1 + P_2\bar{y}_2 + P_3\bar{y}_3$$

$$Var(\bar{y}) = var\left(\sum_{i=1}^{3} p_i y_i\right)$$

$$= \sum_{i=1}^{3} p_i \, Var(\bar{y})$$

Principio da estratificação

Particionar a população de modo que as unidades dentro de cada estrato sejam o mais homogêneas possível

Estrato:

Aula 11/09

Amostragem Estratificada

→ Diminui a variância

N = tamanho população

H = número de estratos

N_h = tamanho do estrato h $(h = 1, 2, \ldots)$

Estimar o parâmetro μ a partir de uma amostra estratificada (AE)

$$\bar{y}_{es} = \sum_{h=1}^{H} w_h \cdot \bar{y}_h \quad \text{onde} \quad w_h = \frac{N_h}{N}$$

$$E(\bar{y}_{es}) = \sum_{h=1}^{H} w_h \, E(\bar{y}_h) \to \mu_h$$

$$= \sum_{h=1}^{H} w_h \cdot \mu_h = \mu$$

entender

$Var(\bar{y}_{es})$ = entender → $(AAS)_c = ✱$

→ $(AAAS)_s = ✱$

$$⊛ \sum_{h=1}^{H} W_h^2 \frac{\sigma^2_h}{n_h}$$

$$⊛ \sum_{h=1}^{H} W_h^2 (1-f_h) \frac{s^2_h}{n_h}$$

$$f_h = \frac{n_h}{N_h}$$

Teorema do Limite Central

$$\frac{\bar{y}_{es} - \mu}{\sqrt{\sum_{h=1}^{H} W_h^2 \, var(\bar{y}_h)}} \sim N(0,1)$$

Alocação das amostras por estrato

Vamos supor que o tamanho da amostra seja n

Importante

→ Alocação proporcional

$$n_h = n \cdot W_h = n \cdot \frac{N_h}{N}$$

AE + AAS_c + Alocação Proporcional

$$var(\bar{y}_{es}) = \sum_{h=1}^{H} W_h^2 \cdot var(\bar{y}_h)$$

$$\overset{AAS_c}{=} \sum_{h=1}^{H} W_h^2 \cdot \frac{\sigma_h^2}{n_h} \overset{AP}{=}$$

$$= \sum_{h=1}^{H} W_h^2 \cdot \frac{\sigma_h^2}{n W_h} = \sum_{h=1}^{H} W_h \frac{\sigma_h^2}{n}$$

Exercicios livros

2.6 $\mathcal{N} = \{1, 2, \ldots, 6\}$

$D = (2, 6, 10, 8, 10, 12)$ → Vetor características populacionais

amostra $n = 2$ é selecionada

sem reposição $\Leftrightarrow f_i(s) \sim \text{Bernoulli}\left(\frac{n}{N}\right)$

a)

① Vamos calcular o numero total de amostras

$f_i(s) \sim \text{Bernoulli}\left(\frac{1}{3}\right)$

$$\binom{N}{n} = \frac{6!}{2!(6-2)!} = \frac{6 \cdot 5 \cdot \cancel{4!}}{2! \cdot \cancel{4!}} = 15$$

Cada amostra possui a mesma probabilidade de seleção $\frac{1}{15}$

$$Var(f_i(s)) = p(1-p)$$
$$= \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$$

⇒ Portanto todas amostras são equiprováveis

Plano Amostral Simétrico

Plano Amostral Simétrico / Plano Equiprovavel

é um tipo de plano de amostragem onde todas as amostras de tamanho n tem a mesma probabilidade de serem selecionadas

Não há vies na seleção

b) $t(s)$ = total da amostra

Encontre a dist. $t(s)$

$$t(s) = \sum_{i \in s} y_i =$$

τ = total população

$$E(t(s)) = E\left(\sum_{i \in s} y_i\right) = E\left[\sum_{i=1}^{N} y_i \, f_i(s)\right]$$

$$= E(f_i(s)) \cdot \tau$$

$$= \frac{n}{N} \cdot \tau^{\mu} = n \cdot \mu$$

$$E(t(s)) = \frac{1}{3} \cdot 48$$

$$= 16$$

$$\frac{n}{N} \cdot \tau = \frac{2}{6} \cdot 48$$

c) Vamos verificar se a média amostral é um estimador não viciado de μ. Calcule $var(\bar{y})$

$$E(\bar{y}) = E\left(\frac{1}{n} \cdot \sum_{i \in s} y_i\right) \qquad f_i(s) \sim \text{Bernoulli}(1/3)$$

$$= E\left(\frac{1}{n} \cdot \underbrace{\sum_{i=1}^{N} y_i}_{\tau} \cdot f_i(s)\right)$$

$$= \frac{1}{n} \cdot \tau \cdot E(f_i(s)) \qquad \tau = 48$$

$$= \frac{1}{2} \cdot 48 \cdot \frac{1}{3} = 8 = \mu$$

$$var(\bar{y}) = var\left(\frac{1}{n} \cdot \underbrace{\sum_{i=1}^{N} y_i}_{\tau} f_i(s)\right) = \frac{\sum (y_i - \bar{y})^2}{N}$$

$$= \frac{1}{n} \cdot \tau^2 \cdot \cancel{var(f_i(s))} \qquad S^2 = \frac{\sum (y - \mu)^2}{N-1}$$

$$\sum (y_i - \mu)^2 = \frac{64}{5} = 12{,}8$$

$$V(\bar{y}) = \frac{1 - f}{n} S^2 \qquad f = \frac{n}{N}$$

3.1 $N = 6$

a) $D = (8, 2, 2, 11, 4, 7)$

Um plano (AAS)s é adotado com $n = 2$ # total Amostras $= \binom{6}{2} = 15$

a) Encontre a dist. $\bar{y}$ e Mostre que $E(\bar{y}) = \mu$

Bern$(\frac{1}{3})$

$$E\left(\frac{1}{n} \cdot \sum_{i=1}^{N} y_i f_i(s)\right) = \frac{1}{2} \cdot \tau \cdot E[f_i(s)]$$

$$\frac{1}{2} \cdot 34 \cdot \frac{1}{3} = \frac{34}{6} = \mu = \text{média populacional}$$

b) $\text{var}(\bar{y}) = \frac{1}{n} \cdot \left(1 - \frac{n}{N}\right) \cdot S^2$

$$= \frac{1}{3} S^2$$

$$S^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (y_i - E(\bar{y}))^2$$

Made with Goodnotes

$$\mathrm{Var}(\bar{y}) = \frac{1}{n} \cdot \left(1 - \frac{n}{N}\right) S^2$$

$$S^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (y_i - E(y))^2$$

3.2 → Refazer o 3.1 com o plano amostral (AAS)c

a) $f_i \sim \mathrm{Binomial}(n, 1/n)$

$$E(\bar{y}) = E\left(\frac{1}{n} \cdot \sum_{i=1}^{N} y_i \cdot f_i(s)\right)$$