

Capítulo 3 Visão Geral da Amostragem e Estimação

3.1 Definições e notação para população de pesquisa e parâmetros selecionados

Nesta seção introduzimos algumas definições e notação necessárias para a apresentação da teoria da amostragem ao longo do texto.

Chama-se *população* (daqui por diante, esta é a designação da *população de pesquisa* que será objeto do levantamento de dados) qualquer conjunto contendo um número finito N de unidades, delimitada por compartilharem algumas características em comum. As unidades deste conjunto são denominadas *unidades da população*. Representamos nosso cadastro dessa população por um conjunto de N rótulos distintos denotado $U = \{1, 2, \dots, i, \dots, N\}$, sendo N o *tamanho da população* e i o rótulo para uma unidade genérica da população. Cada unidade da população fica devidamente identificada por seu rótulo no conjunto U .

São exemplos comuns de populações sobre as quais se realizam pesquisas: domicílios e moradores de certa localidade; indústrias instaladas num certo país; fazendas situadas num certo estado; alunos matriculados na 3^a série do ensino médio da rede escolar estadual em 2017, etc.

No capítulo anterior já enfatizamos a importância de uma definição clara e precisa da *população de pesquisa*. No entanto, ao estudar *Amostragem*, o maior interesse está voltado para o problema de estimar ou inferir certas quantidades ou parâmetros de diversas características (variáveis) numéricas que podem ser medidas ou observadas para cada unidade da população. No caso de características ou variáveis categóricas, podem ser criadas variáveis numéricas indicadoras das categorias de resposta, tomando valor igual a *um* se a unidade é classificada na categoria em questão, e valor igual a *zero* caso contrário. Desta forma, toda a teoria de amostragem se resume à estimação de parâmetros ou quantidades descritivas de variáveis numéricas que poderiam, em tese, ser medidas para todas as unidades da população de pesquisa.

De fato, cada característica numérica ou variável de interesse dá origem a um *vetor populacional*, que é o conjunto de valores da variável correspondentes às unidades da população. Por exemplo, se y é a variável de pesquisa (de interesse) e y_i é o valor dessa variável y para a unidade i , então $Y_U = \{y_1, y_2, \dots, y_i, \dots, y_N\}$ é o *vetor populacional* gerado pela variável y .

Pelo exposto, fica claro que a observação de várias variáveis sobre uma mesma população gera diversos vetores populacionais, cada um correspondendo a uma das variáveis observadas.

Em muitos casos, o interesse em estudar determinada população resume-se à necessidade de conhecer os valores de alguns *parâmetros* de uma ou mais variáveis que podem ser medidas ou observadas para unidades daquela população. Esses *parâmetros-alvo* (ou de interesse) podem ser quaisquer funções dos valores dos vetores populacionais. Os casos mais comuns ocorrem quando há interesse em estimar *totais*, *médias*, *proporções*, *razões*, *quantis* ou mesmo *variâncias*, *covariâncias* e *correlações*, sendo menos frequentes os casos de interesse por outros parâmetros.

Os principais *parâmetros* de interesse podem ser representados por meio das seguintes funções dos valores de variáveis na população.

O *total populacional* da variável y , que é definido por:

$$Y = \sum_{i=1}^N y_i = \sum_{i \in U} y_i$$

A *média populacional* da variável y , que é definida por:

$$\bar{Y} = \frac{Y}{N} = \frac{1}{N} \sum_{i \in U} y_i$$

Uma *proporção populacional* p é simplesmente a média populacional de uma variável y do tipo indicadora, que toma apenas os valores um ou zero para cada unidade. O total de uma variável desse tipo representa a contagem de unidades na população possuidoras do atributo de interesse, e a média é exatamente essa contagem dividida pelo tamanho da população.

Um caso especial de proporção populacional de interesse ocorre com a definição da *função de distribuição cumulativa empírica populacional (FDCEP)*, dada por:

$$F_y(a) = \frac{1}{N} \sum_{i \in U} I(y_i \leq a)$$

onde $a \in \mathbb{R}$ e $I(y_i \leq a)$ representa a função indicadora do evento $y_i \leq a$. A função $F_y(a)$ retorna a proporção de unidades na população U que têm valores de y menores ou iguais a a .

O *quantil populacional* q da distribuição da variável y é definido como o menor valor de y tal que a *FDCEP* tem valor maior ou igual a q , isto é:

$$T_y(q) = \operatorname{argmin}\{F_y(a) \geq q\}$$

onde a função *argmin* retorna o menor valor do argumento a da função $F_y(a)$ tal que a condição $F_y(a) \geq q$ é satisfeita. Por exemplo, a *mediana populacional* da variável y corresponde ao quantil obtido quando $q = 0,5$, isto é, $\text{Mediana}_y = T_y(0,5)$.

A *variância populacional* da variável y é dada por:

$$S_y^2 = \frac{1}{N-1} \sum_{i \in U} (y_i - \bar{Y})^2 = \frac{1}{N-1} \left[\sum_{i \in U} y_i^2 - N\bar{Y}^2 \right]$$

O *desvio padrão populacional* da variável y é dado por:

$$DP_y = S_y = \sqrt{S_y^2}$$

O *coeficiente de variação populacional* da variável y é dado por:

$$CV_y = \frac{DP_y}{\bar{Y}} = \frac{S_y}{\bar{Y}}$$

Seja z outra variável de pesquisa, tomando valores z_i , $i \in U$. Define-se então a *razão de totais* das variáveis y e z como:

$$R = \frac{\sum_{i \in U} y_i}{\sum_{i \in U} z_i} = Y/Z$$

Define-se também a *covariância populacional* e a *correlação populacional* das variáveis y e z como:

$$S_{yz} = \frac{1}{N-1} \sum_{i \in U} (y_i - \bar{Y})(z_i - \bar{Z}) = \frac{1}{N-1} \left[\sum_{i \in U} y_i z_i - N\bar{Y}\bar{Z} \right]$$

e

$$\rho_{yz} = \frac{S_{yz}}{S_y S_z}$$

Até agora foram apresentadas as definições de alguns *parâmetros da população* que se deseja conhecer. No entanto, para conhecer exatamente o valor de qualquer dos parâmetros definidos, seria necessário conhecer **todos** os valores da variável (ou variáveis) de pesquisa naquela população, o que só seria possível mediante a realização de um *Censo* no qual a(s) variável(is) de pesquisa fosse(m) medida(s) ou observada(s) para todas as unidades da população.

Por outro lado, pode ser que estimativas desses parâmetros, com margens de erro conhecidas e controladas, sirvam para os propósitos dos interessados. Nesse caso, uma *pesquisa por amostragem* poderia resolver o problema com vantagens em relação a um *Censo*. Entre as vantagens mais diretas de pesquisas por amostragem podem ser mencionados os menores custos de obtenção das informações de interesse, a maior rapidez para obtenção dos resultados, e a redução da carga de coleta de informações sobre a população de pesquisa.

De agora em diante, consideramos o caso em que a situação enfrentada é tal que basta conhecer *estimativas* dos parâmetros de interesse, bem como indicações da margem de erro a que tais estimativas estão sujeitas. A seguir, tomando por base a ideia de obter *estimativas* dos parâmetros de interesse, são apresentados os principais conceitos relacionados com a *amostragem de populações finitas*.

3.2 Amostra

Uma *amostra* $s = \{i_1, i_2, \dots, i_n\}$ é qualquer subconjunto não vazio de unidades selecionadas da população $U (s \subset U)$ para observação visando *estimar* os parâmetros de interesse. Uma amostra de tamanho n é uma amostra contendo n *unidades* selecionadas da população U , sendo $1 \leq n \leq N$.

A notação $i \in s$ designa que a unidade i foi incluída na amostra s . A notação $s \ni i$ indica que a amostra s contém a unidade i . Quando escrevemos $\sum_{i \in s}$ estamos somando em i sobre o conjunto de rótulos de unidades incluídas na amostra s . Quando escrevemos $\sum_{s \ni i}$ estamos somando em s sobre o conjunto de amostras possíveis que contêm a unidade populacional i .

No contexto deste livro, apresentaremos somente a teoria e resultados aplicáveis a *amostras probabilísticas*, isto é, a amostras selecionadas com base em regras de aleatorização bem definidas e que satisfazem as condições 1 a 3 enunciadas na seção 2.3, e descritas de maneira mais formal na próxima seção.

Os dados amostrais para a variável y serão representados por $Y_s = \{y_{i_1}, y_{i_2}, \dots, y_{i_n}\}$.

3.3 Amostragem probabilística

A *amostragem probabilística* é qualquer procedimento de amostragem que satisfaça todas as condições enumeradas a seguir:

1. O *espaço amostral* S , correspondente ao conjunto de todas as amostras s possíveis, é bem definido, e poderia ser enumerado, ao menos teoricamente.
2. Uma probabilidade $p(s)$ conhecida (ou calculável) é associada a cada amostra $s \in S$, de tal modo que $\sum_{s \in S} p(s) = 1$. A função $p(s)$ é denominada *plano amostral*.
3. Uma única amostra s ($s \in S$) é selecionada para observação usando um mecanismo de aleatorização (sorteio) tal que a amostra s é escolhida com probabilidade igual a $p(s)$.
4. Cada unidade $i \in U$ tem uma probabilidade positiva de ser selecionada para a amostra, isto é: $\pi_i = P(i \in s) = \sum_{s \ni i} p(s) > 0, \forall i \in U$. A probabilidade π_i é denominada *probabilidade de inclusão* (de primeira ordem) da unidade i .
5. As probabilidades de inclusão das unidades selecionadas para a amostra e outros aspectos da estrutura do plano amostral são levados em conta ao fazer inferência sobre os parâmetros populacionais.

3.4 Estatísticas, estimadores e estimativas

Uma *estatística* é uma função real dos valores observados numa amostra da população, isto é, é qualquer função real $f(y_{i_1}, y_{i_2}, \dots, y_{i_n})$. Considere os dois exemplos a seguir:

o *total amostral* ou *soma amostral* da variável y :

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

a *média amostral* da variável y :

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

Um *estimador* $\hat{\theta}(s)$ é uma estatística usada para estimar um certo parâmetro θ de interesse. Antes de observarmos a amostra s , um estimador é uma variável aleatória cuja distribuição temos interesse de conhecer, pois dela dependem propriedades importantes do estimador.

Por simplicidade, daqui para a frente vamos usar a notação $\hat{\theta}$ para designar estimadores, sem explicitar sua dependência da determinação da amostra s , sempre que isso for possível.

Após a determinação da amostra s e a coleta dos dados das unidades nessa amostra, o valor calculado (observado) do estimador é chamado de *estimativa* do parâmetro.

Uma questão central da teoria da amostragem é como escolher bons estimadores para os parâmetros de interesse. Intuitivamente, bons estimadores seriam estatísticas cujos valores fiquem próximos do valor do parâmetro que buscam estimar. Para ajudar com essa questão, é essencial dispor de critérios para a escolha de estimadores. Mas antes disso, é útil definir algumas propriedades de estimadores que serão consideradas na elaboração de critérios de decisão ou escolha que vamos propor usar.

O *valor esperado* de um estimador $\hat{\theta}$ é denotado por $E_p(\hat{\theta})$. A notação $E_p(\cdot)$ designa o valor esperado da quantidade sob a distribuição de probabilidades induzida pelo plano amostral, isto é:

$$E_p(\hat{\theta}) = \sum_{s \in S} \hat{\theta}(s)p(s)$$

O *vício* (ou *viés* ou *tendência*) do estimador $\hat{\theta}$ é definido como:

$$B_p(\hat{\theta}) = E_p(\hat{\theta}) - \theta$$

Algumas vezes é de interesse expressar o vício em termos relativos, e se utiliza então o *vício relativo* do estimador, definido como:

$$RB_p(\hat{\theta}) = \frac{B_p(\hat{\theta})}{\theta}$$

Vício é uma característica indesejada num estimador, pois significa que a distribuição do estimador $\hat{\theta}$ não é centrada no alvo de inferência θ . Diz-se que o estimador $\hat{\theta}$ é *não viciado* (ou *não enviesado* ou *não tendencioso*) para o parâmetro θ quando seu valor esperado é igual ao parâmetro θ , isto é, quando $E_p(\hat{\theta}) = \theta$, ou alternativamente, quando $B_p(\hat{\theta}) = RB_p(\hat{\theta}) = 0$.

Nosso primeiro critério para apoiar a escolha de estimadores sugere então que tratemos de usar *estimadores* sem vício, ou *não viciados*, ou ao menos *aproximadamente não viciados*. Quando isto for possível, teremos estimadores cuja distribuição será centrada no alvo desejado da inferência.

A *variância* do estimador $\hat{\theta}$ é definida como:

$$V_p(\hat{\theta}) = \sum_{s \in S} [\hat{\theta}(s) - E_p(\hat{\theta})]^2 p(s)$$

Quando um estimador é *não viciado*, sua *variância* mede a dispersão da distribuição do estimador em torno do alvo de inferência θ . Duas medidas alternativas dessa dispersão que dependem da variância são o *desvio padrão* ou *DP* do estimador (também designado *erro padrão*), dado por:

$$DP_p(\hat{\theta}) = [V_p(\hat{\theta})]^{1/2}$$

e o *coeficiente de variação* ou *CV* do estimador, dado por:

$$CV_p(\hat{\theta}) = \frac{DP_p(\hat{\theta})}{\theta}$$

O *desvio padrão* mede a dispersão da distribuição do estimador em unidade de medida igual à usada na mensuração da variável de interesse, e o *CV* expressa essa medida em termos relativos, o que pode facilitar a interpretação e a comparação em cenários onde as unidades de medida de diferentes parâmetros podem ser distintas, mas exista interesse em comparar a dispersão de estimadores desses parâmetros.

Quando um estimador $\hat{\theta}$ é *viciado*, uma medida mais adequada da dispersão da distribuição do estimador em torno do alvo de inferência θ é o *erro quadrático médio* ou *EQM* dado por:

$$EQM_p(\hat{\theta}) = \sum_{s \in S} [\hat{\theta}(s) - \theta]^2 p(s)$$

Versões análogas do desvio padrão e do coeficiente de variação adequadas ao caso de estimadores viciados são o *erro médio* ou *EM* e o *erro relativo médio* ou *ERM* definidos, respectivamente, como:

$$EM_p(\hat{\theta}) = [EQM_p(\hat{\theta})]^{1/2}$$

e

$$ERM_p(\hat{\theta}) = \frac{EM_p(\hat{\theta})}{\theta}$$

Um mesmo parâmetro pode ter mais de um estimador não viciado disponível. Precisamos então de um segundo critério para ajudar na escolha de estimadores. Nosso segundo critério vai usar a *variância*, no caso de estimadores exatamente não viciados, ou o *EQM* nos outros

casos. Como se quer ter estimadores com os menores erros de estimação, o segundo critério é o de escolher sempre os estimadores com o menor EQM, ou com a menor variância quando forem não viciados.

No contexto da Amostragem, diferente do contexto usual da Inferência Estatística, não se estabelece uma distribuição de probabilidade (ou modelo) para os valores da variável y na amostra (ou na população). Além disso, em geral os parâmetros que se deseja estimar não são responsáveis pela especificação de uma tal distribuição de probabilidades (ou modelo). Como já indicado, em geral os parâmetros de interesse são definidos como funções dos valores (considerados fixos, mas desconhecidos) da variável y na população.

Por esse motivo, na Amostragem de populações finitas, não há um procedimento geral para gerar estimadores que sejam ótimos nalgum sentido, como é o caso do *método da máxima verossimilhança* no contexto usual da Inferência Estatística. Os princípios usados em Amostragem para derivar estimadores dos parâmetros de interesse são baseados na simplicidade e no *método dos momentos*, como vamos ilustrar.

Suponha que o parâmetro-alvo é o *total populacional* Y . Nesse caso, o objetivo principal seria usar os *dados amostrais* $\{y_{i_1}, y_{i_2}, \dots, y_{i_n}\}$ para *estimar* $Y = \sum_{i \in U} y_i$. Um segundo objetivo seria conseguir medir ou estimar também a *precisão* ou a *margem de erro* da estimativa produzida para Y .

Um *estimador linear* \hat{Y}_w do total populacional Y é uma combinação linear dos valores amostrais y_i com *pesos amostrais* w_i a serem definidos, isto é:

$$\hat{Y}_w = \sum_{i \in s} w_i y_i$$

Podemos então usar os critérios sugeridos para escolha de estimadores para determinar os pesos w_i , como veremos mais adiante.

Para ajudar a consolidar as ideias já apresentadas até aqui, fazemos agora uso de um exemplo muito simples, mas através do qual podemos ilustrar como operar com os conceitos e definições já introduzidos.

Exemplo 3.1 Considere os dados da Tabela 3.1 de uma população fictícia com ($N = 4$) mulheres (unidades populacionais), de quem foi indagado o número de filhos tidos nascidos vivos (a nossa variável y).

Tabela 3.1: Valor da variável y por unidade da população de mulheres

Rótulo da unidade (i)	1	2	3	4	Total
Valor da variável (y_i)	0	0	2	1	3

Existem $\binom{4}{2} = 6$ amostras possíveis de duas unidades distintas dessa população, isto é, de tamanho $n = 2$. O conjunto de todas as amostras possíveis é dado por:
 $S = \{(1; 2); (1; 3); (1; 4); (2; 3); (2; 4); (3; 4)\}.$

A notação para representar o conjunto que forma cada amostra foi o $()$, para evitar usar $\{ \}$ dentro de $\{ \}$. Cada elemento do conjunto S é, em si mesmo, um conjunto (neste caso, um par) de rótulos de unidades selecionadas para a amostra.

Considere agora um *plano amostral* p_1 em que uma qualquer das amostras possíveis é selecionada com igual probabilidade atribuída a todas as amostras possíveis. Considerando a condição 2, cada uma das seis amostras possíveis terá probabilidade igual a $1/6$ de ser selecionada, isto é:

$$p_1(s) = 1/6 \quad \forall s \in S.$$

A Tabela 3.2 apresenta o conjunto de todas as amostras possíveis, os rótulos das unidades incluídas em cada amostra, os valores de y para as unidades incluídas na amostra, a soma amostral e as probabilidades de seleção de cada amostra. As colunas 1, 2 e 5 dessa tabela correspondem à apresentação detalhada do *plano amostral* p_1 tal como definido acima, agora representado na forma de uma tabela.

Tabela 3.2: Informações de cada amostra possível sob plano amostral p_1

Amostra	Unidades na Amostra s	Valores na Amostra s	Soma Amostral (t)	Probabilidades $p_1(s)$
1	{1;2}	{0;0}	0	1/6
2	{1;3}	{0;2}	2	1/6
3	{1;4}	{0;1}	1	1/6
4	{2;3}	{0;2}	2	1/6
5	{2;4}	{0;1}	1	1/6
6	{3;4}	{2;1}	3	1/6
Total	—	—	—	1

A distribuição de probabilidades da estatística *Soma Amostral*, apresentada na Tabela 3.3, pode ser calculada a partir das informações na Tabela 3.2, e é dada por:

Tabela 3.3: Probabilidade sob p_1 para cada valor de t

Valores possíveis de t	0	1	2	3
Com probabilidade $p_1(s)$	1/6	2/6	2/6	1/6

O valor esperado de t é:

$$E_{p_1}(t) = \sum_{s \in S} t(s)p_1(s) = 0 \times \frac{1}{6} + 1 \times \frac{2}{6} + 2 \times \frac{2}{6} + 3 \times \frac{1}{6} = \frac{9}{6} = 1,5$$

Porém o *total populacional* é $Y = \sum_{i \in U} y_i = 3$.

Como $1,5 = E_{p_1}(t) \neq Y = 3$, dizemos que t seria um *estimador viciado* de Y sob o plano amostral p_1 adotado.

Como poderíamos “corrigir” o estimador de modo que ficasse *não viciado* para o total populacional? Uma ideia simples vem da constatação de que $Y/E_{p_1}(t) = 3/1,5 = 2$. Logo, multiplicando por 2 o valor da soma amostral t resultaria num estimador cujo valor esperado deve ser igual a Y .

Considere então o novo estimador do total populacional dado por: $\hat{Y} = 2 \times t$. Tal estimador pode ser escrito na forma linear como: $\hat{Y} = 2 \times t = \sum_{i \in s} 2 \times y_i = \hat{Y}_w$.

A Tabela 3.4 apresenta os valores possíveis e a distribuição de probabilidades do novo estimador $\hat{Y}_w = 2 \times t$.

Tabela 3.4: Probabilidade sob p_1 para cada valor de $2 \times t$

Valores possíveis de $2 \times t$	0	2	4	6
Com probabilidade $p_1(s)$	1/6	2/6	2/6	1/6

Verifica-se então que o valor esperado de $\hat{Y}_w = 2 \times t$ é:

$$E_{p_1}(\hat{Y}_w) = \sum_{s \in S} \hat{Y}_w(s) p_1(s) = 0 \times \frac{1}{6} + 2 \times \frac{2}{6} + 4 \times \frac{2}{6} + 6 \times \frac{1}{6} = \frac{18}{6} = 3$$

Como agora $E_{p_1}(\hat{Y}_w) = 3 = Y$, dizemos que $\hat{Y}_w = 2 \times t$ é um *estimador não viciado* de Y sob o plano amostral p_1 considerado.

O método pelo qual deduzimos os pesos para usar com o estimador \hat{Y}_w não é viável na prática, pois foi preciso conhecer todos os valores da variável de pesquisa para obter o valor esperado do estimador inicialmente considerado (soma amostral), e só então calcular pesos que levariam à obtenção do estimador ponderado não viciado. Essa limitação é resolvida na próxima seção, onde apresentamos um método geral para obter pesos amostrais que levam sempre à estimação não viciada do total populacional.

3.5 A distribuição de aleatorização

A função $p(s)$ definida no conjunto S de todas as amostras possíveis é uma distribuição de probabilidades. Induzida por esta distribuição, é possível obter a distribuição de probabilidades de qualquer estatística (ou estimador) que seria calculada a partir dos dados coletados na amostra selecionada s . A distribuição de probabilidades assim obtida é chamada de *distribuição de aleatorização* da estatística ou estimador. Este foi o conceito ilustrado quando obtivemos a distribuição de probabilidades da estatística *soma amostral* no Exemplo 3.1.

Na *amostragem probabilística*, inferências são feitas considerando a *distribuição de aleatorização*. Tais inferências consideram como única fonte de variação ou incerteza a possível repetição hipotética do processo de amostragem utilizando o *plano amostral* $p(s)$, que resultaria em diferentes amostras $s_1, s_2, \dots \in S$.

A distribuição de $\hat{Y}_w = 2 \times t = \sum_{i \in s} 2 \times y_i$ determinada por $p(s)$ é também chamada de

distribuição amostral do estimador. Vamos estudar suas propriedades para avaliar se \hat{Y}_w é um bom estimador para o total populacional Y .

3.6 Estimadores não viciados para o total populacional

No Exemplo 3.1 mostrou-se como se pode obter a *distribuição amostral* de um estimador (ou de uma estatística qualquer) a partir da distribuição de probabilidades induzida pelo plano amostral $p(s)$. Isto foi muito fácil de fazer porque contamos com duas condições favoráveis, que não se repetirão na prática: a) os tamanhos da população N e da amostra n eram muito pequenos (4 e 2 respectivamente); b) consideramos conhecidos os valores da variável y para todas as unidades da população U .

Na grande maioria das situações de interesse prático no campo das pesquisas por amostragem, os tamanhos de população e amostra serão muito maiores. Também não serão conhecidos os valores da variável de interesse para unidades que não sejam selecionadas para a amostra que vai ser pesquisada. Num cenário com estas características, trabalhar com a distribuição $p(s)$ para daí tentar derivar distribuições amostrais de estimadores é complicado.

O primeiro problema é que o número total de amostras possíveis cresce muito rapidamente com N e com n . Por exemplo, o número de amostras sem reposição de tamanho n de uma população com N unidades é $\binom{N}{n}$. A Tabela 3.5 mostra como cresce o número de amostras no conjunto S para valores selecionados de N e n . Note como o tamanho desse conjunto é gigantesco mesmo com tamanhos de população e amostra bem modestos (1.000 e 20), por exemplo.

Tabela 3.5: Tamanhos do espaço amostral S para valores selecionados de N e n

N	n	$\binom{N}{n}$
4	2	6
10	4	210
100	10	17310309456440
1.000	20	3.39482811302458e+41
10.000	100	6.52084692454763e+241

Uma saída é usar propriedades simplificadoras da distribuição induzida pelo plano amostral. Veremos como fazer isso na próxima seção, mas antes disso, vamos usar uma propriedade importante que pode ser deduzida a partir da distribuição de aleatorização.

Uma propriedade importante

A *probabilidade de inclusão* da unidade i na amostra é dada por:

$$P(i \in s) = \pi_i = \sum_{s \ni i} p(s)$$

Se tomarmos o *inverso da probabilidade de inclusão* $1/\pi_i$ como peso (w_i) de uma unidade amostrada, é fácil verificar que o estimador dado por \hat{Y}_w é *não viciado* para o total populacional Y :

$$\hat{Y}_w = \sum_{i \in s} w_i y_i = \sum_{i \in s} \frac{1}{\pi_i} y_i = \sum_{i \in s} \pi_i^{-1} y_i$$

Essa propriedade é demonstrada de maneira formal na próxima seção. Mas antes disso, vamos verificar sua aplicação com os dados do Exemplo 3.1. Continuando a discussão desse exemplo com a população de ($N = 4$) mulheres de quem foi indagado o número de filhos tidos nascidos vivos (y), tem-se, na Tabela 3.6, o valor da variável y e a probabilidade de inclusão π_i de cada unidade da população de mulheres.

Tabela 3.6: Valor da variável y e probabilidade de inclusão para cada unidade da população

Rótulo da unidade (i)	1	2	3	4	Total
Valor y_i	0	0	2	1	3
Probabilidade de inclusão π_i	3/6=1/2	3/6=1/2	3/6=1/2	3/6=1/2	-

Usando a propriedade recém apresentada, os pesos amostrais no Exemplo 3.1 seriam dados por $w_i = 1/\pi_i = \frac{1}{1/2} = 2$ para qualquer uma das unidades da população que fossem selecionadas para uma das amostras de tamanho $n = 2$.

O estimador ponderado do total nesse caso seria dado por:

$$\hat{Y}_w = \sum_{i \in s} w_i y_i = \sum_{i \in s} \pi_i^{-1} y_i = \sum_{i \in s} 2y_i = 2t$$

e já se mostrou que este estimador é não viciado para Y .

Exemplo 3.2 Considere a mesma população fictícia do exemplo anterior. Considere agora o plano amostral p_2 , que retira amostras de tamanho $n = 2$ - dessa população com as probabilidades indicadas na Tabela 3.7.

Tabela 3.7: Informações de cada amostra possível sob plano amostral p_2

Amostra	Unidades na Amostra s	Valores na Amostra s	Soma Amostral (t)	Probabilidades $p_2(s)$
1	{1;2}	{0;0}	0	0,00
2	{1;3}	{0;2}	2	0,20
3	{1;4}	{0;1}	1	0,15
4	{2;3}	{0;2}	2	0,20
5	{2;4}	{0;1}	1	0,15
6	{3;4}	{2;1}	3	0,30
Total	—	—	—	1,00

Vamos agora usar as informações acima para:

1. Verificar que a estatística soma amostral (t) é viciada para estimar o total populacional Y ;
2. Obter / definir um estimador não viciado para o total populacional Y .

A distribuição da soma amostral t sob o plano p_2 é apresentada na Tabela 3.8.

Tabela 3.8: Probabilidade sob p_2 para cada valor de t

Valores possíveis de t	0	1	2	3
Com probabilidade $p_2(s)$	0,0	0,3	0,4	0,3

O valor esperado de t sob o plano amostral p_2 é:

$$E_{p_2}(t) = \sum_{s \in S} t(s) p_2(s) = 0 \times 0,0 + 1 \times 0,3 + 2 \times 0,4 + 3 \times 0,3 = 2 < 3 = Y$$

Para obter um estimador não viciado, devemos calcular pesos adequados para as unidades amostrais. Estes requerem calcular as probabilidades de inclusão na amostra. A seguir, são apresentadas na Tabela 3.9 as probabilidades de inclusão de cada uma das unidades da população, e também os pesos amostrais correspondentes sob o plano p_2 .

Tabela 3.9: Probabilidade de inclusão e peso amostral de cada unidade sob o plano amostral

p_2

Rótulo da unidade (i)	1	2	3	4
Probabilidade de inclusão (π_i)	7/20=0,35	7/20=0,35	7/10=0,70	3/5=0,60
Peso (w_i)	20/7=2,857	20/7=2,857	10/7=1,429	5/3=1,667

Usando o estimador do total com os pesos adequados \hat{Y}_w , obtém-se os valores das estimativas para cada amostra possível na coluna cinco da Tabela 3.10.

Tabela 3.10: Obtenção de estimativa sob plano amostral p_2 para cada amostra possível

Amostra	Valores na Amostra s	Total Amostral ponderado	Probabilidades $p_2(s)$	Total x probabilidade
1	{0;0}	0	0,00	0
2	{0;2}	2x(10/7)	0,20	4/7
3	{0;1}	1x(5/3)	0,15	1/4
4	{0;2}	2x(10/7)	0,20	4/7
5	{0;1}	1x(5/3)	0,15	1/4
6	{2;1}	2x(10/7)+1x(5/3)	0,30	6/7+1/2
Total	—		1,00	3

Notas

1. O estimador \hat{Y}_w obtido usando os pesos iguais a $1/\pi_i$ tem valor esperado (ver total da última coluna da Tabela 3.10) igual ao total populacional Y . Logo, o estimador assim obtido é *não viciado* também sob o plano amostral p_2 .

2. O fato de que a amostra 1 (composta pelas unidades $\{1;2\}$) tem probabilidade nula de ser selecionada não viola os critérios definidos para que o plano amostral p_2 seja chamado de *amostragem probabilística*. É fácil verificar que todas as condições enumeradas para que uma amostra seja declarada probabilística são cumpridas para esse plano amostral. Em particular, verifica-se que todas as unidades populacionais têm probabilidades positivas de inclusão na amostra - ver a Tabela 3.9.
3. Temos agora duas opções de plano amostral para selecionar amostras (de tamanho $n = 2$) da população U , visando estimar o total populacional Y . Com ambos os planos amostrais estão disponíveis estimadores não viciados do total populacional. Coloca-se então a pergunta: qual dos dois planos é melhor?

Estratégia 1: Seleção equiprovável de amostras com estimador de total ponderado ($\hat{Y}_w = 2t$), conforme especificado na Tabela 3.11.

Tabela 3.11: Probabilidade de seleção sob $p_1(s)$ para cada valor do estimador ponderado

Valores possíveis de $\hat{Y}_w = 2t$	0	2	4	6
Com probabilidade $p(s)$ sob $p_1(s)$	1/6	2/6	2/6	1/6

Estratégia 2: Seleção de amostras com probabilidades desiguais, e estimador de total ponderado (\hat{Y}_w), conforme especificado na Tabela 3.12.

Tabela 3.12: Probabilidade de seleção sob $p_2(s)$ para cada valor do estimador ponderado

Valores possíveis de \hat{Y}_w	5/3	20/7	20/7+5/3
Com probabilidade $p(s)$ sob $p_2(s)$	0,30	0,40	0,30

A melhor estratégia é escolhida medindo o *afastamento esperado* entre os valores possíveis do estimador e o valor do total populacional desconhecido (Y). Para isso, como em ambos os casos o estimador é não viciado, usamos a *variância do estimador*. A Tabela 3.13 indica como pode ser calculada a variância de cada um dos estimadores sob as duas opções de plano amostral (p_1 e p_2).

Tabela 3.13: Obtenção da variância dos estimadores sob os planos amostrais p_1 e p_2

Amostra	Valores	Estimativa sob p_2	Probabilidade sob p_2	Estimativa sob p_1	Probabilidade sob p_1
	na Amostra s				
1	{0;0}	0	0,00	0	1/6
2	{0;2}	$2x(10/7)$	0,20	4	1/6
3	{0;1}	$1x(5/3)$	0,15	2	1/6
4	{0;2}	$2x(10/7)$	0,20	4	1/6
5	{0;1}	$1x(5/3)$	0,15	2	1/6
6	{2;1}	$2x(10/7)+1x(5/3)$	0,30	6	1/6
Variância	—	1,24	—	3,67	—

Conclusão

O plano amostral p_2 fornece o *estimador não viciado com menor variância* em comparação com o plano p_1 , e deve ser preferido, pois o tamanho das amostras (nossa medida de custo) é o mesmo.

Minimizar a variância é o critério de desempate para escolha entre *estratégias não viciadas de amostragem e estimação de igual custo total*. Este será então nosso segundo critério para escolha de estimadores.

3.7 Teoria básica

Nesta seção, seguimos de perto a notação e a forma de apresentar os resultados encontrada no excelente livro de Särndal, Swensson, e Wretman (1992). Outra referência importante é o livro de Fuller (2009).

Como já foi dito, trabalhar com a distribuição $p(s)$ é complicado. Isto ocorre porque o número total $\binom{N}{n}$ de amostras possíveis no conjunto S cresce muito rapidamente com N e com n . A saída encontrada é trabalhar com distribuições das variáveis aleatórias indicadoras $\delta_1, \delta_2, \dots, \delta_N$ definidas tal que:

$$\delta_i = I(i \in s) = \begin{cases} 1 & i \in s \\ 0 & i \notin s \end{cases} \quad \forall i \in U$$

A variável δ_i é indicadora do evento ‘inclusão da unidade i na amostra s ’.

Exemplo 3.1 (Continuação) Para $N = 4$ e $n = 2$, as seis amostras possíveis podem ser representadas pelas indicadoras conforme apresentado na Tabela 3.14.

Tabela 3.14: Representação de cada amostra possível pelas indicadoras

Amostra	Unidades na Amostra s	δ_1	δ_2	δ_3	δ_4
1	{1;2}	1	1	0	0
2	{1;3}	1	0	1	0
3	{1;4}	1	0	0	1
4	{2;3}	0	1	1	0
5	{2;4}	0	1	0	1
6	{3;4}	0	0	1	1

Cada amostra fica univocamente determinada pelas variáveis indicadoras $\delta_1, \delta_2, \dots, \delta_N$ correspondentes. As variáveis indicadoras dependem da amostra s , apesar de não termos indicado isto explicitamente em nossa notação.

As probabilidades de inclusão na amostra, denotadas π_i , podem ser vistas como:

$$\pi_i = P(i \in s) = \sum_{s \ni i} p(s) = P(\delta_i = 1) = E_p(\delta_i) \quad \forall i \in U$$

As probabilidades de inclusão π_i são ditas de *primeira ordem*.

Precisamos também definir *probabilidades de inclusão de segunda ordem*, denotadas π_{ij} , dadas por:

$$\pi_{ij} = P[(i, j) \in s] = \sum_{s \ni (i, j)} p(s) = P(\delta_i \delta_j = 1) = E_p(\delta_i \delta_j) \quad \forall (i, j) \in U$$

Note que quando $i = j$, $\pi_{ij} = \pi_{ii} = \pi_i \quad \forall i \in U$.

Além da propriedade de valor esperado das variáveis aleatórias indicadoras δ_i , pode-se também deduzir que:

$$V_p(\delta_i) = \pi_i(1 - \pi_i)$$

$$COV_p(\delta_i, \delta_j) = \pi_{ij} - \pi_i\pi_j$$

Um Método Geral de Prova em Amostragem

Este método se baseia num uso inteligente das variáveis indicadoras $\delta_1, \delta_2, \dots, \delta_N$. Uma propriedade importante dessas variáveis indicadoras é que:

$$\sum_{i \in s} \delta_i = \sum_{i \in U} \delta_i$$

Segue também que

$$\sum_{i \in s} y_i = \sum_{i \in s} \delta_i y_i = \sum_{i \in U} \delta_i y_i$$

Note que o truque é converter a soma amostral que, antes de selecionada a amostra, tem parcelas aleatórias, em uma soma na população, onde as parcelas são conhecidas mas dependem das variáveis aleatórias indicadoras δ_i .

3.7.1 Estimador linear de total

Considere que o total populacional $Y = \sum_{i \in U} y_i$ é o parâmetro alvo. Um *estimador linear* de Y é sempre da forma:

$$\hat{Y}_w = \sum_{i \in s} w_i y_i = \sum_{i \in U} \delta_i w_i y_i$$

onde w_i é o *peso amostral* da unidade i .

Para que o estimador linear \hat{Y}_w de Y seja **sempre** não viciado, é preciso que:

$$E_p(\hat{Y}_w) = Y \Leftrightarrow \sum_{i \in U} E_p(\delta_i) w_i y_i = \sum_{i \in U} y_i \Leftrightarrow \sum_{i \in U} \pi_i w_i y_i = \sum_{i \in U} y_i$$

Esta relação só será válida para quaisquer valores populacionais y_i da variável de pesquisa caso $\pi_i \times w_i = 1 \quad \forall i \in U$.

Portanto a condição para que o estimador linear do total $\hat{Y}_w = \sum_{i \in s} w_i y_i$ seja **sempre** não

viciado é que os pesos amostrais das unidades selecionadas sejam iguais ao inverso das respectivas probabilidades de inclusão de primeira ordem, isto é:

$$w_i = \pi_i^{-1} = 1/\pi_i = d_i \quad \forall i \in U.$$

Os pesos amostrais d_i são chamados de *pesos básicos* do plano amostral. Doravante usamos a notação d_i para os pesos básicos, pois mais adiante vamos introduzir outros pesos amostrais. A notação w_i é reservada para designar pesos genéricos que podem ser aplicados para a obtenção de estimadores (viciados ou não).

Com os pesos básicos d_i , o estimador não viciado de total fica dado pelo conhecido *estimador de Horvitz-Thompson* ou *estimador HT*:

$$\hat{Y}_{HT} = \sum_{i \in s} d_i y_i = \sum_{i \in s} \pi_i^{-1} y_i = \sum_{i \in s} y_i / \pi_i$$

Este estimador foi proposto por Horvitz e Thompson (1952) e está definido para qualquer variável de pesquisa y e para qualquer *plano amostral probabilístico*, isto é, plano em que $\pi_i > 0 \quad \forall i \in U$. É para permitir desfrutar dessa vantagem de sempre dispor de ao menos um estimador não viciado para totais que esta é uma das condições necessárias para a *amostragem probabilística* de populações finitas. Note também que o estimador faz uso das probabilidades de inclusão implicadas pelo plano amostral $p(s)$ adotado, mas depende deste apenas através das probabilidades de inclusão de primeira ordem das unidades selecionadas para a amostra, uma condição geralmente simples de satisfazer na prática da pesquisa.

3.7.2 Propriedades do estimador de Horvitz-Thompson (#propHT)

O *estimador de Horvitz-Thompson* é *não viciado* para estimar o total, ou seja, $E_p(\hat{Y}_{HT}) = Y$.

Prova:

$$E_p(\hat{Y}_{HT}) = E_p \left[\sum_{i \in U} \delta_i y_i / \pi_i \right] = \sum_{i \in U} \left[E_p(\delta_i) y_i / \pi_i \right] = \sum_{i \in U} y_i = Y$$

Esta propriedade vale para qualquer população, variável de interesse y e plano amostral, desde que $\pi_i > 0 \quad \forall i \in U$.

Variância do estimador Horvitz-Thompson para o total

$$\begin{aligned} V_{HT}(\hat{Y}_{HT}) &= \sum_{i \in U} \sum_{j \in U} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) y_i y_j \\ &= \sum_{i \in U} \sum_{j \in U} \left(\frac{d_i d_j}{d_{ij}} - 1 \right) y_i y_j \end{aligned}$$

onde $d_{ij} = 1/\pi_{ij}$.

Esta é a chamada forma de Horvitz-Thompson da variância. Existe uma outra forma para esta variância, que vamos conhecer mais adiante.

Prova:

$$\begin{aligned}
 V_{HT}(\hat{Y}_{HT}) &= V_p \left(\sum_{i \in U} \delta_i \frac{1}{\pi_i} y_i \right) \\
 &= \sum_{i \in U} \sum_{j \in U} COV_p(\delta_i, \delta_j) \left(\frac{y_i}{\pi_i} \right) \left(\frac{y_j}{\pi_j} \right) \\
 &= \sum_{i \in U} \sum_{j \in U} (\pi_{ij} - \pi_i \pi_j) \left(\frac{y_i}{\pi_i} \frac{y_j}{\pi_j} \right) \\
 &= \sum_{i \in U} \sum_{j \in U} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) y_i y_j \\
 &= \sum_{i \in U} \sum_{j \in U} \left(\frac{d_i d_j}{d_{ij}} - 1 \right) y_i y_j
 \end{aligned}$$

Estimador da variância do estimador de total

Um estimador não viciado da variância do estimador HT do total é dado por:

$$\hat{V}_{HT}(\hat{Y}_{HT}) = \sum_{i \in s} \sum_{j \in s} (d_i d_j - d_{ij}) y_i y_j$$

Este estimador da variância foi obtido usando o princípio dos estimadores tipo Horvitz-Thompson do total, mas agora, como se tratava de estimar uma soma dupla na população, os pesos das parcelas nessa soma dependem das probabilidades de inclusão de segunda ordem, isto é, das probabilidades de inclusão dos pares de unidades. Para que este estimador seja viável, o plano amostral empregado tem que satisfazer a condição adicional de que as probabilidades de inclusão π_{ij} sejam estritamente positivas $\forall i \neq j \in U$.

Forma alternativa para a variância do estimador HT do total

Para planos amostrais de tamanho pré-fixado, pode-se demonstrar que uma forma equivalente da variância do estimador de Horvitz-Thompson do total populacional é dada pela expressão de Sen-Yates-Grundy a seguir - ver Yates e Grundy (1953) e Sen (1953).

$$\begin{aligned}
 V_{SYG}(\hat{Y}_{HT}) &= \sum_{i \in U} \sum_{j > i} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \\
 &= \sum_{i \in U} \sum_{j > i} (1/d_i d_j - 1/d_{ij}) (d_i y_i - d_j y_j)^2
 \end{aligned}$$

Note a troca do sinal da diferença de probabilidades de inclusão em relação à expressão anterior.

Uma análise dessa expressão de variância nos dá uma indicação de quando empregar probabilidades de inclusão distintas pode ser vantajoso. A variância do estimador de total

seria nula caso $\frac{y_i}{\pi_i} = \frac{y_j}{\pi_j} \forall i \neq j \in U$. Isto só seria possível quando $\pi_i \propto y_i \forall i \in U$, isto é,

quando as probabilidades de inclusão fossem exatamente proporcionais aos valores da variável de interesse. Na prática, é impossível aplicar essa ideia já que os valores da variável de interesse são desconhecidos antes da seleção da amostra.

Entretanto, veremos no capítulo 10 que esta ideia pode ser usada de forma aproximada fazendo as probabilidades de inclusão proporcionais a uma medida de tamanho cujos valores estejam disponíveis para todas as unidades da população U . Sempre que a medida de tamanho for positivamente correlacionada com a(s) variável(is) de interesse y , veremos que é possível tirar proveito da informação de tamanho para aplicar métodos de amostragem que levam a estimadores mais eficientes do total que no caso de planos amostrais com equiprobabilidade para amostras de tamanhos iguais.

Estimador alternativo da variância do estimador HT do total

$$\begin{aligned}
 \hat{V}_{SYG}(\hat{Y}_{HT}) &= \sum_{i \in s} \sum_{j > i} \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \right) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \\
 &= \sum_{i \in s} \sum_{j > i} (d_{ij}/d_i d_j - 1) (d_i y_i - d_j y_j)^2
 \end{aligned}$$

O estimador $\hat{V}_{SYG}(\hat{Y}_{HT})$ foi motivado a partir da forma de Sen-Yates-Grundy para a variância do estimador HT do total. Tal estimador não coincide com o estimador de variância derivado a partir da expressão de Horvitz-Thompson apresentada anteriormente.

Comentários finais

1. Com amostras probabilísticas, é sempre possível estimar sem vício um total populacional usando uma soma amostral π -ponderada, isto é, o estimador HT do total.

2. Há expressões de variância disponíveis para permitir avaliar a qualidade do estimador de total sob distintas situações (população, variável) para qualquer plano amostral probabilístico.
3. A estimação de muitos outros parâmetros populacionais, tais como médias, proporções, razões, etc. usa em grande medida os resultados aqui apresentados para a estimação de totais. Isto ficará mais claro nos capítulos seguintes.
4. Pode-se derivar estimadores não viciados do total populacional, e da variância do estimador HT de total para distintos planos amostrais como casos especiais da teoria geral aqui apresentada. Isto será conveniente, em particular, para a estimação de variâncias, cujas expressões gerais dependem de somas duplas que podem se tornar inconvenientes de calcular quando os tamanhos de amostra são grandes. As expressões apresentadas para cada um dos planos amostrais específicos são úteis porque permitem simplificar os cálculos da estimação de variâncias.

3.8 Exercícios

Exercício 3.1 Considere a população com $N = 6$ domicílios listada com os respectivos valores de variáveis de interesse na Tabela 3.15.

Tabela 3.15: Valores de variáveis de interesse para cada domicílio da população

Domicílio	Renda (R\$)	Número de Moradores	Número de Trabalhadores
1	800	2	2
2	4.200	4	3
3	1.600	2	1
4	500	2	1
5	900	4	2
6	2.000	1	1
Total	10.000	15	10

Tratar das seguintes questões:

1. Para cada variável de interesse (Renda, Número de Moradores e Número de Trabalhadores), calcule os seguintes parâmetros populacionais:

- a. Total;
 - b. Média;
 - c. Variância.
2. Liste o conjunto S de todas as amostras possíveis de tamanho 2 da população, considerando apenas amostras de unidades distintas.
 3. Supondo que todas as amostras listadas no conjunto S são equiprováveis (Plano A), calcule:
 - a. As probabilidades de inclusão das unidades;
 - b. As probabilidades de inclusão dos pares de unidades;
 - c. Os valores possíveis para o estimador Horvitz-Thompson do total populacional para a variável Renda;
 - d. O valor esperado e a variância para o estimador Horvitz-Thompson do total populacional para a variável Renda.
 4. Considere agora que o conjunto S é formado somente pelas amostras (1;2), (2;3), (2;4), (2;5) e (2;6), tendo cada uma delas probabilidade 1/5 de ser a amostra selecionada (Plano B). Repita a) a d) do item 3 para este novo plano amostral.
 5. Faça gráficos dos valores possíveis do estimador de total sob os dois planos amostrais para comparar as respectivas distribuições.
 6. Use os resultados obtidos em 3 e 4 para comparar os dois planos amostrais e indique qual deles seria preferível usar, caso fosse necessário amostrar duas unidades distintas da população ($n = 2$) para estimar o total da *Renda*. Justifique.