3.3 Amostragem aleatória simples sem reposição

i. A população está numerada de 1 a N, de acordo com o sistema de referências, ou seja,

$$\mathcal{U} = \{1, \ldots, N\};$$

ii. Utilizando uma tabela de números aleatórios, ou programa de computador, sorteia-se, com igual probabilidade, uma das N unidades da população;

iii. Sorteia-se um elemento seguinte, com o elemento anterior sendo retirado da população.

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

Teorema 3.7 *Com relação à AASs, a variável f_i, número de vezes que a unidade i aparece na amostra, obedece a distribuição de Bernoulli (ver Bussab e Morettin, 2004) com probabilidade de sucesso n/N, denotado por $f_i \sim b(1; n/N)$, e que satisfaz:*

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

de modo que

$$E[f_i] = \frac{n}{N},$$

$$Var[f_i] = \frac{n}{N}\left(1 - \frac{n}{N}\right),$$

$$\pi_i = \frac{n}{N},$$

$$\pi_{ij} = \frac{n}{N}\frac{n-1}{N-1},$$

e

$$Cov[f_i, f_j] = -\frac{n}{N^2}\frac{N-n}{N-1},$$

$i = 1, \ldots, n$ *e* $i \neq j = 1, \ldots, N$.

3.3.1 Propriedades da estatística $t(\mathbf{s})$

Apresenta-se a seguir algumas propriedades da estatística $t(\mathbf{s}) = \sum_{i\in\mathbf{s}} Y_i$, o total da amostra, que serão utilizadas na seção seguinte, quando são apresentados estimadores do total e da média populacionais e suas propriedades.

Teorema 3.8 *Com relação à AASs, a estatística* $t(\mathbf{s})$ *tem as seguintes propriedades:*

$$E[t] = n\mu$$

e

$$Var[t] = n(1-f)S^2,$$

onde $f = n/N$ *é denominada fração amostral.*

3.3.2 Estimação do total e da média populacional

Corolário 3.3 *Com relação à AASs, um estimador não viciado do total populacional é*

$$T(\mathbf{s}) = N\overline{y} = \frac{N}{n}t(\mathbf{s}),$$

cuja variância amostral é dada por

$$Var[T] = N^2(1-f)\frac{S^2}{n}.$$

Corolário 3.4 *Com relação à AASs, a média amostral*

$$\overline{y} = \frac{1}{n}\sum_{i\in\mathbf{s}}^{n} Y_i = \frac{t(\mathbf{s})}{n}$$

é um estimador não viesado da média populacional, com variância amostral dada por

$$Var[\overline{y}] = (1-f)\frac{S^2}{n}.$$

3.3.3 Estimação da variância populacional

Teorema 3.9 *A variância da amostra*

$$s^2 = \frac{1}{n-1} \sum_{i \in \mathbf{s}} (Y_i - \overline{y})^2$$

é um estimador não viesado da variância populacional S^2 *para o planejamento AASs.*

Corolário 3.5 *Para o plano amostral AASs, a estatística*

$$var[\overline{y}] = \widehat{Var}[\overline{y}] = (1-f)\frac{s^2}{n}$$

é um estimador não viesado de $Var[\overline{y}]$ *e*

$$var[T] = \widehat{Var}[T] = N^2(1-f)\frac{s^2}{n}$$

é um estimador não viesado de $Var[T]$*.*

Exemplo 3.4 Considere novamente os dados do Exemplo 2.1 e o interesse pela variável renda familiar, onde, como no Exemplo 3.1, $\mathcal{U} = \{1, 2, 3\}$ e o parâmetro populacional é $\mathbf{D} = (12, 30, 18)$, com as funções paramétricas $\tau = 60$, $\mu = 20$ e $S^2 = 84$. Definido o plano amostral AASs, com $n = 2$, tem-se associado a $\mathcal{U}$ o espaço amostral

$$\mathcal{S}_{\text{AASs}} = \{12, 21, 31, 13, 23, 32\}.$$

A Tabela 3.4 considerada a seguir apresenta os valores de $\overline{y}$ e s^2 para cada uma das amostras em $\mathcal{S}_{\text{AASs}}$.

Tabela 3.4: Valores de $\overline{y}$, s^2 e $P(\mathbf{s})$ para as amostras $\mathbf{s}$ em $\mathcal{S}_{\text{AASs}}$

$\mathbf{s}$:	12	21	13	31	23	32
$P(\mathbf{s})$:	1/6	1/6	1/6	1/6	1/6	1/6
$\overline{y}$:	21	21	15	15	24	24
s^2:	162	162	18	18	72	72

As Tabelas 3.5 e 3.6 apresentam as distribuições amostrais de $\overline{y}$ e s^2.

Tabela 3.5: Distribuição amostral de $\overline{y}$ na AASs

$\overline{y}$:	15	21	24
$P(\overline{y})$:	1/3	1/3	1/3

Temos da Tabela 3.5 que

$$E[\overline{y}] = 20 \quad \text{e} \quad Var[\overline{y}] = \left(1 - \frac{2}{3}\right)\frac{84}{2} = 14.$$

3.3.4 Normalidade assintótica e intervalos de confiança

Tabela 3.6: Distribuição amostral de s^2 na AASs

s^2:	18	72	162
$P(s^2)$:	1/3	1/3	1/3

Portanto, $\overline{y}$ é um estimador não viesado para μ e com variância bem menor que a variância apresentada pelo planejamento AASc. Da Tabela 3.6, tem-se que

$$E[s^2] = \frac{1}{3}(18 + 72 + 162) = 84,$$

um resultado já esperado, pois, conforme visto no Teorema 3.9, s^2 é um estimador não viesado de S^2.

Exemplo 3.5 Uma pesquisa amostral foi conduzida com o objetivo de se estudar o índice de ausência ao trabalho em um determinado tipo de indústria. Uma AAS sem reposição de mil operários de um total de 36 mil é observada com relação ao número de faltas não justificadas em um período de 6 meses. Os resultados obtidos foram:

Faltas:	0	1	2	3	4	5	6	7	8
Trabalhadores:	451	162	187	112	49	21	5	11	2

Para esta amostra tem-se que uma estimativa de μ é dada por $\overline{y} = 1,296$. Observa-se também que $s^2 = 2,397$. Usando a aproximação normal, tem-se que um intervalo de 95% de confiança para μ é dado por $(1,201; 1,391)$.

3.3.5 Determinação do tamanho da amostra

Exemplo 3.6 Considere a população dos operários faltosos do Exemplo 3.5. Pode-se encontrar n tal que $B = 0,05$, com $\alpha = 0,05$. Então, como, neste caso $D \cong (0,05/2)^2 = 0,00065$, tem-se que

$$n \cong \frac{1}{\frac{0{,}00065}{2{,}397} + \frac{1}{36.000}} \cong 3.466.$$

Pode-se também considerar o caso em que o interesse é pelo erro máximo relativo como no caso da AASc considerado na Seção 3.2.5. Veja também o Exercício 3.42.

3.3.6 Estimação de proporções

Exemplo 3.7 No Exemplo 3.5, suponha que até 3 faltas (3 dias) em 6 meses seja considerado aceitável. Então, a proporção de trabalhadores tirando mais que 3 dias de folga não justificada em 6 meses, é

$$p = \frac{88}{1000} = 0,088.$$

De (3.34), tem-se que um intervalo de confiança para P, com $\alpha = 0,05$, é dado por

$$0,088 \pm 1,96\sqrt{\left(1 - \frac{1.000}{36.000}\right)\frac{0,088 \times 0,912}{1.000 - 1}},$$

ou seja, $(0,071; 0,105)$. Para ter uma estimativa com $B = 0,01$ com $\gamma = 0,95$, temos que $D \cong (0,01/2)^2 = 0,000025$, de modo que de (3.34) segue que é preciso observar

$$n = \frac{36.000}{\frac{(36.000-1)\times 0,000025}{0,088\times 0,912} + 1} \cong 2.948.$$