

# Capítulo 8 Amostragem Sistemática Simples (AS)

## 8.1 Ideia básica

A Amostragem Sistemática Simples (AS) é um método para selecionar com *equiprobabilidade* unidades de uma população de pesquisa para compor a amostra desejada. Por sua simplicidade, trata-se de uma alternativa à AAS bastante utilizada na prática. Um dos motivos para justificar sua utilização é que o método pode ser aplicado mesmo quando não existe cadastro prévio da população de pesquisa, podendo este cadastro ser construído ao mesmo tempo em que é feita a seleção da amostra.

O método consiste em selecionar cada  $K$ -ésima unidade da população a ser amostrada, começando de uma *partida aleatória*  $r$  sorteada ao acaso entre os números inteiros de 1 a  $K$ . A amostra resultante é *sem reposição*, e todas as unidades da população têm a mesma chance ( $1/K$ ) de serem selecionadas, como se verá adiante.

Seja  $U = \{1, 2, \dots, N\}$  a população de pesquisa, de tamanho  $N = nK + c$ , com  $0 \leq c < K$ . O número inteiro  $K$  é chamado de *intervalo de seleção*,  $n = \lfloor N/K \rfloor$  é igual à parte inteira da divisão de  $N$  por  $K$  e  $c$  é o resto dessa mesma divisão.

AS pode ser utilizada para selecionar uma amostra de um cadastro de unidades da população mediante um algoritmo de processamento sequencial de lista. Na ausência de um cadastro prévio, o método também pode ser implementado, bastando para isso especificar um mecanismo para cadastramento sequencial das unidades populacionais e definir o valor de  $K$ . Isto feito, deve ser selecionada a *partida aleatória* e iniciado o cadastramento sequencial das unidades da população tal como especificado. O algoritmo de seleção das unidades da amostra pode ser facilmente implementado em ambos os casos (cadastro prévio ou construído durante a operação de pesquisa).

**Exemplo 8.1** Suponha que se deseja aplicar um questionário a uma amostra dos espectadores de uma peça teatral sendo encenada em determinado teatro, num determinado dia. Nesse caso não estaria disponível uma lista das pessoas que irão ao teatro naquela data. Pode-se selecionar uma AS utilizando os passos indicados a seguir. 1) Definir o valor de  $K$ . Por exemplo, seja  $K = 10$  significando que ao final a amostra será composta por

aproximadamente um de cada dez, ou 10% dos espectadores da noite. 2) Selecionar a *partida aleatória* entre 1 e 10. Suponha que o número escolhido seja 3. 3) Iniciar o processo de cadastramento sequencial dos espectadores de acordo com a ordem de chegada (ou de saída) ao teatro, numerando cada espectador cadastrado. 4) Entrevistar os espectadores selecionados por AS, começando pelo de número 3 e seguindo com os de números de chegada (ou saída) iguais a 13, 23, 33, . . .

Duas características funcionam como um apelo para a adoção da AS como método de seleção da amostra: sua simplicidade, já que ao selecionar a primeira unidade a ser incluída na amostra, todas as demais estarão automaticamente escolhidas; e a possibilidade de aplicação mesmo quando não se tem disponível um cadastro prévio da população.

Essas características foram determinantes, por exemplo, da adoção de AS para a seleção, em cada setor censitário, dos domicílios que deveriam responder ao questionário da amostra nos Censos Demográficos realizados pelo IBGE de 1960 até 2000. Para implementar o método, cada recenseador utilizava um formulário denominado *Folha de coleta do setor* para cadastrar os domicílios que ia encontrando ao percorrer um setor censitário que lhe fora atribuído. Este formulário servia para cadastrar as unidades da população (domicílios) encontradas em cada setor, e de instrumento para a seleção da amostra sistemática. Linhas marcadas em sombreado indicavam ao recenseador em que domicílios devia aplicar o questionário mais longo, denominado da amostra, em lugar do questionário simplificado (denominado básico), que era aplicado aos demais domicílios não incluídos na amostra.

Ao terminar de percorrer o setor censitário, o recenseador teria elaborado um cadastro dos domicílios de seu setor, e selecionado a amostra correspondente. O cadastro de domicílios assim construído era utilizado como base para o trabalho de campo das outras pesquisas por amostragem realizadas pelo IBGE ao longo da década subsequente ao Censo. Veja por exemplo Albieri e Bianchini (2015).

Nos censos de 1960, 1970 e 1980 o intervalo de seleção usado pelo IBGE foi sempre  $K = 4$ . Começando no Censo de 1991, o IBGE passou a usar valores de  $K$  que podiam variar conforme o tamanho do município em que a amostra estava sendo selecionada. Maiores detalhes sobre a amostragem nos Censos de 2000 e 2010 podem ser vistos em IBGE (2003) e IBGE (2016), respectivamente. Uma revisão dos aspectos de amostragem dos Censos Demográficos brasileiros desde 1960 pode ser encontrada em Albieri (2017).

## 8.2 Método de seleção da amostra

O método de seleção de uma *amostra sistemática simples* pode ser generalizado a partir do Exemplo 8.1, consistindo nos seguintes passos:

1. Defina o valor de  $K$ , o *intervalo de seleção* da AS;
2. Selecione a *partida aleatória*  $r$ , igual a um número inteiro sorteado entre 1 e  $K$  com probabilidades iguais a  $1/K$  para todos os inteiros no intervalo; sendo assim,  $r$  tem distribuição Uniforme Discreta de parâmetro  $K$ , ou seja,  $r \sim UD(K)$ ;
3. Inclua na amostra sistemática  $s_r$  todas as unidades que satisfazem a regra indicada abaixo.

$$s_r = \{i : i = r + lK \leq N; \quad l = 0, \dots, n\} \quad (8.1)$$

Em consequência desse método, há exatamente  $K$  *amostras sistemáticas* distintas possíveis. Todas têm igual probabilidade de ser a amostra selecionada, logo:

$$p(s) = \begin{cases} 1/K & \text{se } s = s_r, \quad r = 1, 2, \dots, K \\ 0 & \text{caso contrário} \end{cases} \quad (8.2)$$

O *tamanho efetivo* da amostra selecionada por AS, aqui denotado por  $n_r$ , não é fixado a priori, pois depende do valor da *partida aleatória*  $r$  selecionada, e pode tomar dois valores possíveis:

$$n_r = \begin{cases} n + 1 & \text{quando } r \leq c \\ n & \text{quando } r > c \end{cases}$$

O *tamanho efetivo* da amostra será  $n + 1$  com probabilidade  $c/K$  e  $n$  com probabilidade  $1 - (c/K)$ .

**Exemplo 8.2** Considere uma população de  $N = 20$  unidades, da qual se quer selecionar uma amostra sistemática simples com intervalo de seleção  $K = 5$ . Nesse caso, existem 5 amostras distintas, todas de tamanho  $n = 4$ . Pode-se utilizar o R para realizar a tarefa.

```
# Fixa semente para geração de números aleatórios
set.seed(867)

# Tamanho da população
N=20

print(paste("Tamanho da população N:", N), quote=FALSE)

## [1] Tamanho da população N: 20
```

```
# Intervalo de seleção
K=5
print(paste("Intervalo de seleção K:", K), quote=FALSE)

## [1] Intervalo de seleção K: 5

# Cálculo do valor de n (tamanho desejado da amostra)
n=trunc(N/K)
print(paste("Tamanho da amostra n:", n), quote=FALSE)

## [1] Tamanho da amostra n: 4

# Calculo do valor de c
c=N-n*K
print(paste("Valor de c:", c), quote=FALSE)

## [1] Valor de c: 0

# Geração da partida aleatória
r=sample(1:K,1)
print(paste("Partida aleatória r:", r), quote=FALSE)

## [1] Partida aleatória r: 4

# Geração da lista das unidades da população U
U=1:N
# Determinação das unidades da amostra s
s_r = subset(U, (U%%K)==r)
print("Amostra sistemática selecionada s_r:", quote=FALSE)
```

```
## [1] Amostra sistemática selecionada s_r:
```

```
print(s_r)
```

```
## [1] 4 9 14 19
```

## 8.3 Composição de amostras sistemáticas simples de intervalo $K$

As unidades componentes das possíveis amostras sistemáticas de intervalo  $K$  podem ser esquematicamente representadas como nas linhas da Tabela 8.1 a seguir.

Tabela 8.1: Unidades que compõe as possíveis  $K$  amostras sistemáticas com partida aleatória no intervalo  $[1, 2, \dots, K]$

Amostras possíveis	$l = 0$	$l = 1$	$\vdots$	$l = n$
		$K + 1$		$nK + 1$
$s_1 \ s_2 \ \dots \ s_c \ \dots \ s_K$	$1 \ 2 \ \dots \ c$	$K + 2 \ \dots$	$\dots \dots \dots$	$nK + 2 \ \dots$
	$\dots \ K$	$K + c \ \dots$	$\dots \dots \dots$	$nK + c -$
		$2K$		$-$

Examinando a Tabela 8.1 se conclui que unidades em diferentes amostras sistemáticas simples possíveis nunca podem aparecer juntas na mesma amostra! Outra constatação importante é que o conjunto de amostras possíveis é bem pequeno, e de tamanho  $K$ , em contraste com o que ocorre com os planos de amostragem AAS e AASC, discutidos até agora.

**Exemplo 8.3** Identifique as amostras sistemáticas simples possíveis quando a população tem  $N = 19$  unidades e o tamanho desejado da amostra é de  $n = 4$  unidades.

Como  $N = 19 = 4 \times 4 + 3$ , temos que  $K = 4$  e  $c = 3$ . Logo, as quatro amostras sistemáticas possíveis nesse caso são:

$$s_1 = \{1; 5; 9; 13; 17\} \text{ com } n_1 = 5;$$

$$s_2 = \{2; 6; 10; 14; 18\} \text{ com } n_2 = 5;$$

$s_3 = \{3; 7; 11; 15; 19\}$  com  $n_3 = 5$ ;

$s_4 = \{4; 8; 12; 16\}$  com  $n_4 = 4$ .

Podemos utilizar o R para realizar a mesma tarefa.

```
# Tamanho da População
N=19
print(paste("Tamanho da população N:", N), quote=FALSE)

## [1] Tamanho da população N: 19

# Tamanho desejado da amostra
n=4
print(paste("Tamanho desejado da amostra n:", n), quote=FALSE)

## [1] Tamanho desejado da amostra n: 4

# Cálculo do intervalo de seleção
K=trunc(N/n)
print(paste("Intervalo de seleção K:", K), quote=FALSE)

## [1] Intervalo de seleção K: 4

# Cálculo da constante c
c=N-n*K
print(paste("Constante c:", c), quote=FALSE)

## [1] Constante c: 3
```

```
# Vetor para guardar nomes das amostras possíveis
nomes_amostras = paste0("s_", 1:K)

# Matriz para guardar os índices das unidades das amostras possíveis
amostras_possiveis=matrix(c(1:N,NA),K,n+1,byrow=FALSE)

# Monta arquivo com resultado das amostras possíveis
amostras_possiveis = as.data.frame(cbind(nomes_amostras, amostras_possiveis))
colnames(amostras_possiveis) = c("Amostra", paste("Unidade", 1:(n+1)))
amostras_possiveis
```

```
##      Amostra Unidade 1 Unidade 2 Unidade 3 Unidade 4 Unidade 5
## 1      s_1          1          5          9         13         17
## 2      s_2          2          6         10         14         18
## 3      s_3          3          7         11         15         19
## 4      s_4          4          8         12         16        <NA>
```

**Exemplo 8.4** Calcule o tamanho efetivo da amostra resultante da seleção sistemática em uma população com  $N = 149$  unidades, quando o tamanho desejado da amostra sistemática simples é de  $n = 60$  unidades. Como  $N = 149 = 60 \times 2 + 29$ , resulta que  $K = 2$  com  $n = 74$  e  $c = 1$ . Sendo assim as duas únicas amostras possíveis são:

$s_1 = \{\text{números ímpares até } 149, \text{ inclusive}\}$  com  $n_1 = 75$ ;

$s_2 = \{\text{números pares até } 148, \text{ inclusive}\}$  com  $n_2 = 74$ ;

Nesse caso, verifica-se que o tamanho efetivo da amostra poderá ser 74 ou 75, um pouco maiores que o tamanho desejado de 60.

## 8.4 Probabilidades de inclusão na amostragem sistemática

É fácil notar que a probabilidade de inclusão na amostra de uma unidade  $i$  qualquer é dada por:

$$\pi_i = \frac{1}{K}, \quad i = 1, \dots, N$$

Foi visto na Equação (8.1) que a primeira unidade a ser incluída na amostra é a unidade  $r$ , que é um inteiro selecionado com equiprobabilidade no intervalo  $[1, K]$ . Como há  $K$  números inteiros nesse intervalo, segue-se que a probabilidade de sortear um qualquer desses números é  $1/K$ .

As demais unidades selecionadas são obtidas somando a  $r$  os múltiplos  $lK$ , com  $l$  variando de 1 a  $n$ , enquanto  $lK + r \leq nK + c$ , varrendo assim os intervalos de seleção subsequentes. Desta forma, ao selecionar  $r$  todas as unidades amostrais ficam definidas, implicando que a probabilidade de selecionar uma unidade qualquer da população é igual à de selecionar a partida aleatória que determina qual das amostras  $s_r$  será observada.

A probabilidade de inclusão das unidades  $i \neq j$  na amostra é dada por:

$$\pi_{ij} = \begin{cases} 1/K & \text{se } i \neq j \in s_r, \quad r = 1, \dots, K, \\ 0 & \text{caso contrário.} \end{cases}$$

## 8.5 Variáveis aleatórias indicadoras - propriedades

Note que ao escolher  $r$ ,  $1 \leq r \leq K$ , selecionamos a amostra inteira. Sejam as variáveis indicadoras associadas às amostras possíveis  $s_r$ :

$$I(r) = \begin{cases} 1 & \text{se a amostra é } s_r, \text{ para } 1 \leq r \leq K. \\ 0 & \text{caso contrário.} \end{cases}$$

Então o valor esperado de  $I(r)$  é

$$E_{AS}[I(r)] = 1/K, \quad r = 1, \dots, K,$$

A variância é dada por:

$$V_{AS}[I(r)] = E_{AS}\{[I(r)]^2\} - \{E_{AS}[I(r)]\}^2 = \frac{1}{K} - \frac{1}{K^2} = \frac{1}{K} \left(1 - \frac{1}{K}\right),$$

E a covariância entre  $I(r)$  e  $I(q)$  quando  $r \neq q$  é:

$$COV_{AS}[I(r), I(q)] = E_{AS}[I(r)I(q)] - E_{AS}[I(r)]E_{AS}[I(q)] = 0 - \frac{1}{K^2} = -\frac{1}{K^2}$$

já que apenas uma das duas partidas  $r$  ou  $q$  pode ser selecionada.



## 8.6 Estimação de totais sob amostragem sistemática simples

Para apresentar o estimador tipo Horvitz-Thompson do total sob AS, denotamos por  $t_r$  a soma amostral dos valores observados da variável  $y$  para a amostra  $s_r$ , definida como:

$$t_r = \sum_{i \in s_r} y_i$$

Como a probabilidade de inclusão de uma amostra sistemática simples  $s_r$  qualquer é  $1/K$ , o peso amostral das unidades dessa amostra é sempre igual a  $d_i = 1/\pi_i = K$ . Sendo assim, sob AS o estimador de Horvitz-Thompson para o total é dado por:

$$\hat{Y}_{AS} = K t_r = K \sum_{i \in s_r} y_i \quad (8.3)$$

Com base nas propriedades do estimador de Horvitz-Thompson para o total, já sabemos que este estimador é não viciado para o total populacional. Mas vamos aqui demonstrar esse resultado também para o caso particular da AS, pois a prova nos ajudará com a obtenção posterior de expressão para a variância do estimador sob esse plano amostral. Note então que:

$$\begin{aligned} E_{AS}(\hat{Y}_{AS}) &= E_{AS}[K t_r] = K E_{AS} \sum_{r=1}^K I(r) t_r \\ &= \sum_{r=1}^K E_{AS}[I(r)] t_r = K \sum_{r=1}^K \frac{1}{K} t_r = \sum_{r=1}^K t_r = Y \end{aligned}$$

provando, portanto, que  $\hat{Y}_{AS}$  é não viciado para o total populacional  $Y$ .

Quando o tamanho populacional  $N$  for desconhecido, pode também ser estimado sem vício usando uma variável  $y$  útil para contagens, tomando o valor 1 para todas as unidades da população:

$$\hat{N}_{AS} = K n_r$$

Um estimador do total de unidades da população pode ser útil em situações nas quais não há cadastro prévio, e o mesmo é construído durante a operação de pesquisa.

**Exemplo 8.5** Considere a população composta de  $N = 19$  unidades, cujos dados da variável de interesse  $y$  são mostrados na Tabela 8.2, da qual se deseja retirar uma amostra sistemática simples com intervalo de seleção  $K = 4$  para estimar o total populacional.

Verifique numericamente que o estimador  $\hat{Y}_{AS}$  é não viciado.

Tabela 8.2: Dados das amostras sistemáticas possíveis, numa população com  $N = 19$  e intervalo de seleção  $K = 4$

$s_1$	$s_2$	$s_3$	$s_4$
99	54	96	54
85	88	55	83
62	85	96	55
91	92	67	68
54	79	76	

Para mostrar que o estimador é não viciado, basta verificar que a média dos seus valores possíveis é igual ao parâmetro populacional a ser estimado. No caso, basta calcular as estimativas de total para cada uma das possíveis amostras (cada coluna da Tabela 8.2), e calcular a média dessas estimativas. A comparação com o total populacional, dado pela soma de todos os valores da tabela, mostra o resultado desejado. Pode-se utilizar o R para realizar a tarefa, como ilustrado a seguir.

```
# Intervalo de seleção
```

```
K=4
```

```
print(paste("Intervalo de seleção K:", K), quote=FALSE)
```

```
## [1] Intervalo de seleção K: 4
```

```
# Carregando os valores populacionais da variável y
```

```
pop=matrix(c(99,54,96,54,85,88,55,83,62,85,96,55,91,92,67,68,54,79,76,NA),5,K,byrow=T)
```

```
print("Dados populacionais", quote=FALSE)
```

```
## [1] Dados populacionais
```

```
print(pop)
```

```
##      [,1] [,2] [,3] [,4]
## [1,]  99  54  96  54
## [2,]  85  88  55  83
## [3,]  62  85  96  55
## [4,]  91  92  67  68
## [5,]  54  79  76  NA
```

```
# Calculando a estimativa do total para cada uma das AS possíveis
```

```
Yhatr=NULL
```

```
for(l in (1:K)) Yhatr[l] = K*sum(pop[,l], na.rm=T)
```

```
print("Valores possíveis do estimador de total", quote=FALSE)
```

```
## [1] Valores possíveis do estimador de total
```

```
print(Yhatr)
```

```
## [1] 1564 1592 1560 1040
```

```
# Calculando a média das estimativas de total
```

```
EYhat=mean(Yhatr)
```

```
print(paste("Média das estimativas possíveis do total:", EYhat), quote=FALSE)
```

```
## [1] Média das estimativas possíveis do total: 1439
```

```
# Calculando o total populacional Y
```

```
Y=sum(pop, na.rm=T)
```

```
print(paste("Total populacional Y:", Y), quote=FALSE)
```

```
## [1] Total populacional Y: 1439
```

## 8.7 Estimação de médias sob amostragem sistemática simples

A média populacional  $\bar{Y} = Y/N$  pode ser vista como uma razão entre o total da variável  $y$  e o total de unidades populacionais:

$$\bar{Y} = \frac{Y}{N} = \frac{\sum_{r=1}^K t_r}{\sum_{r=1}^K n_r}$$

Para estimá-la deve-se considerar duas situações: quando o tamanho  $N$  da população é conhecido, ou quando esse valor é desconhecido. Um estimador não viciado para a média quando  $N$  é conhecido é dado por:

$$\bar{y}_{AS} = \frac{\hat{Y}_{AS}}{N} = \frac{K t_r}{N} \quad (8.4)$$

Este estimador é não viciado para  $\bar{Y}$ , pois foi visto que  $\hat{Y}_{AS}$  é não viciado para o total populacional  $Y$ . Também é interessante observar que este estimador não é igual à média amostral, a menos no caso em que  $N = nK$ .

**Exemplo 8.6** Considerando a mesma população do Exemplo 8.5 pode-se verificar que  $\bar{y}_{AS}$  é não viciado para a média populacional  $\bar{Y}$ . Neste caso, basta calcular a estimativa da média para cada coluna (AS possível) da Tabela 8.2, calcular o valor médio das estimativas e comparar com média populacional. Pode-se utilizar o R para realizar a tarefa.

```
# Intervalo de seleção e tamanho da população conhecido
```

```
K=4
```

```
print(paste("Intervalo de seleção K:", K), quote=FALSE)
```

```
## [1] Intervalo de seleção K: 4
```

```
# Carregando os valores populacionais da variável y
```

```
pop=matrix(c(99,54,96,54,85,88,55,83,62,85,96,55,91,92,67,68,54,79,76,NA),5,K,byrow=T)
```

```
print("Dados populacionais", quote=FALSE)
```

```
## [1] Dados populacionais
```

```
print(pop)
```

```
##      [,1] [,2] [,3] [,4]  
## [1,]   99   54   96   54  
## [2,]   85   88   55   83  
## [3,]   62   85   96   55  
## [4,]   91   92   67   68  
## [5,]   54   79   76   NA
```

```
# Calcula tamanho da população
```

```
N=sum(!is.na(pop))
```

```
print(paste("Tamanho da população N:", N), quote=FALSE)
```

```
## [1] Tamanho da população N: 19
```

```
# Calculando a estimativa da média para cada uma das AS possíveis
```

```
ybarAS.r=NULL
```

```
for(i in (1:K)) ybarAS.r[i] = K*sum(pop[,i], na.rm=T)/N
```

```
print("Valores possíveis do estimador não viciado da média", quote=FALSE)
```

```
## [1] Valores possíveis do estimador não viciado da média
```

```
print(ybarAS.r)
```

```
## [1] 82.31579 83.78947 82.10526 54.73684
```

```
# Calculando a média das estimativas
EybarAS=mean(ybarAS.r)
print(paste("Média das estimativas possíveis da média:", EybarAS), quote=FALSE)
```

```
## [1] Média das estimativas possíveis da média: 75.7368421052632
```

```
# Calculando a média populacional
Ybar=mean(pop, na.rm=T)
print(paste("Média populacional:", Ybar), quote=FALSE)
```

```
## [1] Média populacional: 75.7368421052632
```

**Exemplo 8.7** Utilizando a mesma população, pode-se verificar que o estimador média amostral simples ( $\bar{y}$ ) para a média populacional de  $y$  não coincide com o estimador  $\bar{y}_{AS}$ , e além disso, é viciado.

```
# Intervalo de seleção e tamanho da população conhecido
K=4
# Carregando os valores populacionais da variável y
pop=matrix(c(99,54,96,54,85,88,55,83,62,85,96,55,91,92,67,68,54,79,76,NA),5,K,byrow=T)
# Calculando a média amostral simples para cada uma das AS possíveis
ybar.r=NULL
for(i in (1:K)) ybar.r[i] = mean(pop[,i], na.rm=T)
ybar.r
```

```
## [1] 78.2 79.6 78.0 65.0
```

```
# Calculando a média das estimativas
Eybar=mean(ybar.r)
print(paste("Média das estimativas possíveis da média:", Eybar), quote=FALSE)
```

```
## [1] Média das estimativas possíveis da média: 75.2
```

Quando  $N$  é desconhecido, um estimador alternativo para a média é dado pela razão entre os estimadores do total e do tamanho populacionais:

$$\bar{y}_{AS} = \frac{\widehat{Y}_{AS}}{\widehat{N}_{AS}} = \frac{K t_r}{K n_r} = \frac{t_r}{n_r} = \bar{y}_r = \bar{y} \quad (8.5)$$

que é a média amostral.

Verifica-se assim que a média amostral simples é um estimador para uma razão, sendo portanto viciado para estimar a média populacional. Tal estimador só será exatamente não viciado quando  $N = nK$ , pois:

$$\begin{aligned} E_{AS}(\bar{y}_{AS}) &= E_{AS}(\bar{y}) = E_{AS} \sum_{r=1}^K I(r) \bar{y}_r \\ &= \frac{1}{K} \sum_{r=1}^K \bar{y}_r = \frac{1}{K} \sum_{r=1}^K \frac{t_r}{n_r} \\ &\neq \frac{\sum_{r=1}^K t_r}{\sum_{r=1}^K n_r} = \bar{Y} X, \end{aligned}$$

O vício desse estimador (quando  $N \neq nK$ ) é o preço pago quando não se conhece o tamanho  $N$  da população!

## 8.8 Estimação de uma proporção na amostragem Sistemática

Como já foi visto no Capítulo @ref(cap:proporc), muitas vezes o interesse é estimar a proporção de unidades da população que possuem uma determinada característica de interesse. Nesse caso, para as unidades onde a característica de interesse estiver presente a variável  $y$  terá o valor 1 e para as demais unidades  $y$  receberá o valor 0.

Como já foi visto que a proporção equivale à média de uma variável do tipo indicadora, quando o tamanho total da população,  $N$ , for conhecido um estimador não viciado para a proporção é dado por:

$$\hat{p}_{AS} = \frac{K}{N} \sum_{i \in s_r} y_i = \frac{K}{N} t_r = \frac{K}{N} n_a \quad (8.6)$$

onde  $n_a$  denota o número de unidades na amostra que têm a característica de interesse.

**Exemplo 8.8** Utilizando o R pode-se ver numericamente que o estimador  $\hat{p}_{AS}$  para a proporção  $p$  é não viciado quando  $N$  é conhecido.

```
# Intervalo de seleção e tamanho da população conhecido
K=4
print(paste("Intervalo de seleção K:", K), quote=FALSE)

## [1] Intervalo de seleção K: 4

# Carregando os valores populacionais da variável y
(pop=matrix(c(0,1,0,1,1,1,0,0,0,0,1,0,0,0,1,1,1,0,1,NA),5,K,byrow=T))

##      [,1] [,2] [,3] [,4]
## [1,]    0    1    0    1
## [2,]    1    1    0    0
## [3,]    0    0    1    0
## [4,]    0    0    1    1
## [5,]    1    0    1   NA

# Calcula tamanho da população
N=sum(!is.na(pop))
print(paste("Tamanho da população N:", N), quote=FALSE)

## [1] Tamanho da população N: 19

# Calcula estimativas da proporção para cada uma das AS possíveis
p_AS=NULL
for(i in (1:K)) p_AS[i] = K*sum(pop[,i],na.rm=T)/N
print("Valores possíveis do estimador não viciado da proporção", quote=FALSE)

## [1] Valores possíveis do estimador não viciado da proporção
```



```
print(p_AS)

## [1] 0.4210526 0.4210526 0.6315789 0.4210526

# Calculando a média das estimativas
Ep_AS=mean(p_AS)
print(paste("Média das estimativas possíveis da proporção:", Ep_AS), quote=FALSE)

## [1] Média das estimativas possíveis da proporção: 0.473684210526316

# Calculando proporção populacional p
p=mean(pop, na.rm=TRUE)
print(paste("Proporção na população p:", p), quote=FALSE)

## [1] Proporção na população p: 0.473684210526316
```

No caso em que o tamanho  $N$  da população é desconhecido pode-se utilizar o estimador tipo razão que, nesse caso, é igual à proporção amostral:

$$\hat{p}_{AS} = \frac{n_a}{n_r},$$

Este estimador é viciado, a menos do caso onde  $N = nK$ , como já foi visto.

## 8.9 Variância dos estimadores sob amostragem sistemática simples

A variância de  $\hat{Y}_{AS}$  sob amostragem sistemática simples é dada por:

$$\begin{aligned}
V_{AS}(\hat{Y}_{AS}) &= V_{AS} \left[ K \sum_{r=1}^K I(r) t_r \right] \\
&= K^2 \left[ \sum_{r=1}^K t_r^2 V_{AS}[I(r)] + \sum_{r \neq q} COV_{AS}[I(r), I(q)] t_r t_q \right] \\
&= K^2 \left[ \sum_{r=1}^K t_r^2 \frac{1}{K} \left( 1 - \frac{1}{K} \right) + \sum_{r \neq q} t_r t_q \left( -\frac{1}{K^2} \right) \right] \\
&= K^2 \left[ \frac{1}{K} \sum_{r=1}^K t_r^2 - \frac{1}{K^2} \left( \sum_{r=1}^K t_r^2 + \sum_{r \neq q} t_r t_q \right) \right] \\
&= K \left[ \sum_{r=1}^K t_r^2 - \left( \sum_{r=1}^K t_r \right)^2 / K \right] \\
&= K \sum_{r=1}^K (t_r - \bar{t})^2
\end{aligned} \tag{8.7}$$

$$\text{onde: } \bar{t} = \frac{1}{K} \sum_{r=1}^K t_r = \frac{Y}{K}$$

Portanto a variância é calculada a partir da soma de quadrados dos desvios entre totais das amostras possíveis em relação à média destes totais.

Quando  $N$  é conhecido a variância do estimador da média populacional é dada por:

$$V_{AS}(\bar{y}_{AS}) = \frac{1}{N^2} V_{AS}(\hat{Y}_{AS})$$

Um problema que ocorre com a amostragem sistemática simples é o fato da ordenação da população em relação aos valores da variável de interesse afetar a variância (precisão) dos estimadores. Para demonstrar esse problema pode-se utilizar a população apresentada nos exemplos acima, alterando a ordem de suas unidades, para mostrar como isso afeta a composição das possíveis amostras sistemáticas.

Em primeiro lugar, consideramos a população ordenada tal como foi apresentada na Tabela 8.2 e calculamos a variância do estimador do total considerando as possíveis amostras listadas no exemplo 8.5. Em seguida, ordenamos a população em ordem crescente (ou decrescente) dos valores de  $y$  e repetimos o cálculo da variância. Observa-se que a variância do estimador do total com a população ordenada diminui consideravelmente em comparação com o caso em que não havíamos reordenado a população.

Embora este seja um exemplo extremo, pois na prática não teríamos como ordenar a população pelos valores da variável de interesse, ele mostra bem o efeito que a ordenação pode ter na precisão dos estimadores baseados em amostras sistemáticas simples. Na sequência, ilustramos estes cálculos usando o R.

```
# Intervalo de seleção e tamanho da população conhecido
```

```
K=4
```

```
print(paste("Intervalo de seleção K:", K), quote=FALSE)
```

```
## [1] Intervalo de seleção K: 4
```

```
# População na ordem natural
```

```
# Carregando os valores populacionais da variável y
```

```
pop=matrix(c(99,54,96,54,85,88,55,83,62,85,96,55,91,92,67,68,54,79,76,NA),5,K,byrow=T)
```

```
print("Dados populacionais", quote=FALSE)
```

```
## [1] Dados populacionais
```

```
print(pop)
```

```
##      [,1] [,2] [,3] [,4]
```

```
## [1,]  99   54   96   54
```

```
## [2,]  85   88   55   83
```

```
## [3,]  62   85   96   55
```

```
## [4,]  91   92   67   68
```

```
## [5,]  54   79   76   NA
```

```
# Calcula tamanho da população
```

```
N=sum(!is.na(pop))
```

```
# Calculando a estimativa do total para cada uma das AS possíveis
```

```
tr=NULL
```

```
for(i in (1:K)) tr[i] = sum(pop[,i],na.rm=T)
```

```
# Calculando a variância do estimador do total
```

```
V_YhatAS= K*(var(tr)*(K-1))
```

```
print(paste("Variância do estimador de total - ordem natural:", V_YhatAS), quote=FALSE)
```

```
## [1] Variância do estimador de total - ordem natural: 53219
```

```
# Ordenando a população em ordem crescente de y
pop_ord = matrix(sort(pop, na.last=T), 5, K, byrow=T)
print("Dados populacionais ordenados", quote=FALSE)
```

```
## [1] Dados populacionais ordenados
```

```
print(pop_ord)
```

```
##      [,1] [,2] [,3] [,4]
## [1,]  54  54  54  55
## [2,]  55  62  67  68
## [3,]  76  79  83  85
## [4,]  85  88  91  92
## [5,]  96  96  99  NA
```

```
# Calculando a estimativa do total para cada uma das AS possíveis
```

```
tr=NULL
```

```
for(i in (1:K)) tr[i] = sum(pop_ord[,i], na.rm=T)
```

```
# Calculando a variância do estimador do total
```

```
V_YhatAS=K*(var(tr)*(K-1))
```

```
print(paste("Variância do estimador de total - ordem crescente:", V_YhatAS), quote=FALSE)
```

```
## [1] Variância do estimador de total - ordem crescente: 20611
```

Fica como exercício para o leitor verificar que o mesmo não ocorre quando se utiliza uma AAS ou AASC, pois nestes casos a ordenação dos valores de  $y$  na população não tem influência na precisão das estimativas.

## Notas

1. Como  $r$  pode tomar apenas um valor,  $V_{AS}(\hat{Y}_{AS})$  não pode ser diretamente estimada a partir da amostra.
2. Em Cochran (1977) é apresentada uma boa discussão sobre como a ordenação dos valores da variável de pesquisa para unidades populacionais pode afetar a eficiência de amostras sistemáticas.
3. Para populações em ‘ordem aleatória’, o desempenho da amostragem sistemática simples é semelhante ao da amostragem aleatória simples sem reposição (Cochran (1977), seção 8.5).
4. Para populações com tendência linear, amostragem sistemática simples é melhor que AAS (Cochran (1977), seção 8.6).
5. Para populações periódicas, amostragem sistemática simples com intervalo de seleção em sincronia com o período é um desastre (Cochran (1977), p. 218).

**Considere o caso especial onde  $N = n \times K$**

Da tabela 8.2 do Cochran (1977) é fácil ver que  $\bar{y}_r = t_r/n_r$  é um estimador não viciado para  $\bar{Y}$ . Como

$$\sum_{r=1}^K \sum_{i \in s_r} (y_i - \bar{Y})^2 = n \sum_{r=1}^K (\bar{y}_r - \bar{Y})^2 + \sum_{r=1}^K \sum_{i \in s_r} (y_i - \bar{y}_r)^2$$

tem-se que

$$(N - 1)S_y^2 = nKV_{AS}(\bar{y}) + K(n - 1)S_D^2$$

onde  $S_y^2$  é a variância populacional total,  $S_D^2$  é a variância *dentro* das amostras sistemáticas e  $V_{AS}(\bar{y})$  é a variância de  $\bar{y}$  sob amostragem sistemática simples.

Portanto, conforme Cochran (1977), p. 208, o estimador de média é mais eficiente sob amostragem sistemática que sob AAS se e somente se  $S_D^2 > S_y^2$ .

**Uma expressão alternativa para  $V_{AS}(\bar{y})$  quando  $N = nK$**

De acordo com o Teorema 8.2 de Cochran (1977), p. 209, tem-se:

$$\begin{aligned} V_{AS}(\bar{y}) &= \left( \frac{N-1}{N} \right) [1 + (n-1)\rho] \frac{S_y^2}{n} \\ &\doteq [1 + (n-1)\rho] \frac{S_y^2}{n} \end{aligned} \tag{8.8}$$

onde:

$$\rho = \frac{1}{(n-1)(N-1)S_y^2} \sum_{r=1}^K \sum_{i \neq j \in s_r} (y_i - \bar{Y})(y_j - \bar{Y}) \quad (8.9)$$

é a *correlação intraclasse* das amostras sistemáticas possíveis.

Este resultado mostra que a correlação positiva entre unidades de uma mesma amostra aumenta a variância da média amostral sob seleção sistemática simples quando comparada com a variância da média amostral sob seleção AAS.

Lembrando que sob AAS,  $V_{AAS}(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}$ .

Então, tem-se a relação aproximada:

$$\frac{V_{AS}(\bar{y})}{V_{AAS}(\bar{y})} = 1 + (n-1)\rho \quad \begin{array}{ll} < 1 & \text{se } \rho < 0, \\ = 1 & \text{se } \rho = 0, \\ > 1 & \text{se } \rho > 0. \end{array} \quad (8.10)$$

Cabe notar que razões do tipo  $V_{AS}(\bar{y})/V_{AAS}(\bar{y})$  são chamadas de *Efeito do Plano Amostral (EPA)*. Em inglês, usa-se o termo *design effect*, abreviado como *deff*.

## 8.10 Estimação de variâncias dos estimadores na amostragem sistemática

Quando se usa amostragem sistemática simples não há um estimador não viciado para a variância dos estimadores do total e da média. O que se faz é utilizar estimadores mais adequados de acordo com a ordenação da população.

No caso de se poder supor que não há uma ordenação ou que a população esteja espalhada aleatoriamente no cadastro de seleção em relação à(s) variável(eis) de interesse, de modo que as médias dentro de cada intervalo de seleção sejam pelo menos aproximadamente iguais ( $\bar{Y}_r \doteq \text{constante}$ ), o que se pode fazer é utilizar um estimador para a variância da média equivalente ao usado quando se tem uma AAS, dado pela fórmula:

$$\hat{V}_{1AS}(\bar{y}_{AS}) = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{n-1} \sum_{i \in s_r} (y_i - \bar{y}_{AS})^2 \quad (8.11)$$

Esse estimador serve para o caso de  $N$  conhecido e será não viciado caso a suposição de ordenação aleatória das unidades na população esteja correta. No caso de não se conhecer  $N$  pode-se utilizar as alternativas dadas pelos estimadores de razão.

No caso em que a população esteja ordenada segundo uma “estratificação” de modo que as médias em cada intervalo de seleção variem (p.ex.: a população é ordenada segundo os valores de  $y$ ), pode-se utilizar, para estimar a variância do estimador da média, a fórmula:

$$\widehat{V}_{2AS}(\bar{y}_{AS}) = \left( \frac{1}{n} - \frac{1}{N} \right) \frac{1}{2(n-1)} \sum_{i \in s_r} (y_i - y_{i+K})^2 \quad (8.12)$$

Alternativamente e independente da ordenação da população, pode-se usar um estimador do tipo replicação, onde são selecionadas  $q$  amostras sistemáticas de tamanhos  $n/q$  cada uma, tomando a variância das estimativas dadas por cada uma das amostras. Essa técnica, também chamada *amostra sistemática repetida*, está descrita em Scheaffer et al. (2011).

Quando a seleção de uma AS for realizada a partir de um cadastro conhecido, é sempre possível reordenar as unidades aleatoriamente antes proceder a seleção. Esse é um artifício muito útil e que permite que se utilize os estimadores equivalentes aos da AAS para estimar a variância dos estimadores. Se por um lado essa técnica viabiliza o emprego de estimadores simplificados de variância, por outro se espera que acabe resultando em menor precisão para a estimação pontual.

Pode-se encontrar boas discussões sobre a estimação da variância sob amostragem sistemática simples em Cochran (1977) ou Thompson (2012).

A Tabela 8.3 apresenta um resumo sobre a estimação de alguns parâmetros na AS.

Tabela 8.3: Principais resultados para os estimadores sob AS

Parâmetro	Estimador sob AS
$Y = \sum_{i \in U} y_i$	$\hat{Y}_{AS} = K t_r = K \sum_{r=1}^K I(r) t_r$
$N$	$\hat{N}_{AS} = K n_r$ , se $N$ é desconhecido
$\bar{Y} = Y/N$	$\bar{y}_{AS} = \frac{K}{N} t_r$ , se $N$ é desconhecido $\bar{y}_{AS} = \frac{t_r}{n_r} = \bar{y}$ , se $N$ é desconhecido
$p = N_A/N$	$\hat{p}_{AS} = \frac{K}{N} n_a$ , se $N$ é desconhecido
$V_{AS}(\hat{Y}_{AS}) = K \sum_{r=1}^K (t_r - \bar{t})^2$	$\hat{p}_{AS} = \frac{n_a}{n_r}$ , se $N$ é desconhecido Depende da ordenação das unidades na população amostrada
$V_{AS}(\bar{y}_{AS}) = \frac{K}{N^2} \sum_{r=1}^K (t_r - \bar{t})^2$	Depende da ordenação das unidades na população amostrada

## 8.11 Alternativas para seleção com amostragem sistemática simples

Como foi visto, existem dificuldades quando o tamanho da população,  $N$ , não é um múltiplo do intervalo de seleção  $K$ , ou seja, quando  $N \neq nK$ . Neste caso, o estimador simples da média têm vício. Cochran (1977) indica que esse vício pode ser considerado desprezível quando se trabalha com tamanhos de amostra razoavelmente grandes, podendo-se considerar como tal amostras com tamanhos iguais ou maiores que  $n = 50$ .

Há também alternativas que podem ser utilizadas quando a seleção da amostra sistemática for feita a partir de um cadastro ou lista das unidades populacionais. Cochran (1977) recomenda usar um método de seleção ‘circular’ proposto por Lahiri em 1952, para obter uma amostra sistemática. Esse método consiste em: 1) Tomar como intervalo de seleção o inteiro mais próximo de  $N/n$ , ou  $K = \text{round}(N/n)$ ; 2) Selecionar como partida aleatória um número inteiro  $r \in [1, N]$ ; 3) Tomar como primeira unidade da amostra a unidade  $r$ ; 4) Em seguida, selecionar as unidades seguintes sempre somando  $K$  ao índice da última unidade selecionada; quando  $r + jK > N$ , subtrair  $N$  e continuar o processo até obter as  $n$  unidades amostrais desejadas.



**Exemplo 8.9** Seja uma população de  $N = 21$  unidades da qual se deseja uma amostra sistemática de exatamente  $n = 5$  unidades. O *script* a seguir seleciona uma amostra sistemática circular com o tamanho desejado. Note que nesse caso  $K = 4$ .

```
# Definindo a semente para permitir replicação
```

```
set.seed(77099)
```

```
# Tamanho da população
```

```
N=21
```

```
# Tamanho exato da amostra desejada
```

```
n=5
```

```
# Passo 1: calculando o valor de K
```

```
(K=round(N/n))
```

```
## [1] 4
```

```
# Passo 2: selecionando a partida aleatória r
```

```
r=sample(1:N,1)
```

```
# Definindo a primeira unidade amostral
```

```
sr=NULL
```

```
sr[1]=r
```

```
# Definindo as demais unidades amostrais
```

```
for(i in (2:n)){
```

```
  sr[i]=sr[i-1]+K
```

```
  if(sr[i]>N) sr[i]=sr[i]-N
```

```
}
```

```
# Exibindo a partida
```

```
r
```

```
## [1] 10
```

```
# Exibindo a amostra selecionada
```

```
sr
```

```
## [1] 10 14 18 1 5
```

Veja que com a amostra sistemática circular o número de amostras possíveis é  $N$  e pode-se definir estimadores não viciados para a média e o total da variável de interesse  $y$ , como:

$$\bar{y}_{ASc} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad \text{e} \quad \hat{Y} = N\bar{y}_{ASc}$$

Neste método a seleção é feita com equiprobabilidade e sem reposição, como na AS tradicional, porém é necessário que  $N$  seja conhecido e a seleção é um pouco mais trabalhosa. A vantagem é que os estimadores são sempre não viciados.

Fica para o leitor verificar que os estimadores para a média e total são não viciados.

## 8.12 Exercícios

**Exercício 8.1** Identifique as outras 4 amostras possíveis referidas no Exemplo 8.1.

**Exercício 8.2** A Tabela 8.4 mostra a listagem dos moradores dos 13 domicílios de uma rua. As pessoas foram listadas de acordo com a seguinte regra: H=homem adulto, M=mulher adulta, h=homem criança e m=mulher criança. Cada coluna da tabela representa um dos domicílios da rua.

Tabela 8.4: Composição das famílias dos domicílios da rua

Domicílios	1	2	3	4	5	6	7	8	9	10	11	12	13
	H	H	H	H	H	H	H	H	H	H	H	H	H
	M	M	M	M	M	M	M	M	M	M	M	M	M
	m	m	h		h	m	m	h	h	h	m	m	
	h	h	m		h	h	m	m		m	h		
	m	m			m		h						

- Calcule a variância do estimador da proporção de homens adultos usando AS de 1 em cada 5 pessoas;
- Calcule a variância do estimador da proporção de homens adultos usando AAS com fração amostral de 20%;
- Repita os itens *a* e *b* para a proporção de crianças;
- Em que casos a AS é mais indicada que a AAS? Por que?

**Dica:** para selecionar a amostra sistemática numere as pessoas sequencialmente a partir do homem adulto do domicílio 1, descendo na coluna e passando para o topo da coluna seguinte.

**Exercício 8.3** Para estimar o número de erros de digitação num livro de 207 páginas, uma amostra sistemática foi retirada selecionando-se, em primeiro lugar, um número aleatório entre 1 e 10, incluindo na amostra todas as décimas páginas subsequentes. O número aleatório selecionado foi 9. Os números de erros nas páginas selecionadas foram, respectivamente, 1, 1, 3, 0, 0, 0, 2, 3, 0, 2, 2, 1, 1, 2, 2, 2, 2, 2, 3 e 3.

- Quais foram as páginas selecionadas?
- De uma estimativa não viciada do total de erros de digitação no livro.
- Uma pessoa usou os dados da amostra e calculou uma estimativa para o total de erros da seguinte maneira: calculou o número médio de erros por página da amostra e multiplicou por 207, resultando 331,20 erros. Esta é uma estimativa não viciada considerando o plano amostral utilizado? Justifique sua resposta?
- A variância do estimador foi estimada por  $207 \times (207 - 20) \times 1,09/20$ , onde 1,09 é a variância amostral dos 20 valores observados. Essa estimativa é não viciada? Justifique sua resposta?

**Exercício 8.4** A Tabela 8.5 mostra o número de sementes germinadas em cada canteiro de um viveiro de mudas com um total de 200 canteiros. Suponha que se deseja estimar a média de sementes por canteiro, através de uma amostra de tamanho 10. (**Dica:** As linhas,  $i$ , são as possíveis amostras sistemáticas e as colunas,  $h$ , representam uma estratificação dos canteiros)

Tabela 8.5: Número de sementes germinadas por canteiro

<b>i \ h</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>Totais</b>
1	8	20	26	34	31	24	18	16	36	10	223
2	6	19	26	21	23	19	13	12	8	35	182
3	6	25	10	27	41	28	7	8	29	7	188
4	23	11	41	25	18	18	9	10	33	9	197
5	25	31	30	32	15	29	11	12	14	12	211
6	16	26	55	43	21	24	20	20	13	7	245
7	28	29	34	33	8	33	16	17	18	6	222
8	21	19	56	45	22	37	9	12	20	14	255
9	22	17	39	23	11	32	14	7	13	12	190
10	18	28	41	27	3	26	15	17	24	15	214
11	26	16	27	37	4	36	20	21	29	18	234
12	28	9	20	14	5	20	21	26	18	4	165
13	11	22	25	14	11	43	15	16	16	4	177
14	16	26	39	24	9	27	14	18	20	9	202
15	7	17	24	18	25	20	13	11	6	8	149
16	22	39	25	17	16	21	9	19	15	8	191
17	44	21	18	14	13	18	25	27	4	9	193
18	26	14	44	38	22	19	17	29	8	10	227
19	31	40	55	36	18	24	7	31	8	5	255
20	26	30	39	29	9	30	30	29	10	3	235
Totais	410	459	674	551	325	528	303	358	342	205	4155

- Calcule a variância do estimador da média se for utilizada uma amostra sistemática;
- Calcule a variância do estimador da média usando uma AAS de mesmo tamanho;
- Comente os resultados.

**Exercício 8.5** A administração do Parque da Tijuca precisou estimar o número médio de ocupantes nos carros que visitaram o parque num determinado dia. Para isso foi feita uma amostra, usando a seguinte estratégia:

- Na hora da abertura da cabine de cobrança foi selecionado um número aleatório entre 1 e 10;
  - A partir do veículo com esse número, por ordem de chegada, foram contados os ocupantes de todos os décimos veículos.

No final do dia passaram pela cancela 105 veículos e os selecionados para a amostra tinham, por ordem de passagem, 3, 5, 7, 2, 3, 4, 5, 4, 4, e 2, ocupantes (incluindo o motorista).

- a. Defina qual o tipo de plano amostral utilizado e dê os seus parâmetros;
- b. Liste os números de ordem dos veículos pesquisados, sabendo que o número aleatório selecionado foi 8;
- c. Estime a média de ocupantes por veículo no dia da pesquisa;
- d. Supondo que a distribuição do número de ocupantes é aleatória, estime um intervalo de 95% de confiança para a média de ocupantes por veículo.

**Exercício 8.6** O departamento de controle de qualidade de uma indústria alimentícia utiliza amostragem sistemática para estimar a média da quantidade de determinado alimento colocada em cada lata na linha de produção. A Tabela 8.6 relaciona a amostra selecionada de 1 em cada 50 latas num dia, quando a produção diária foi de 1820 latas.

Tabela 8.6: Quantidade de alimento, em onças, por lata na AS.

12,00	11,97	12,01	12,03	12,01	11,80
11,91	11,98	12,03	11,98	12,00	11,83
11,87	12,01	11,98	11,87	11,90	11,88
11,75	11,93	11,95	11,97	11,93	12,05
11,85	11,98	11,87	12,05	12,02	12,04
12,05	11,87	11,91	11,93	11,94	11,89

- a. Estime a média da quantidade de alimento por lata;
- b. Estime a variância da estimativa da média e de um intervalo de 95% de confiança para a média;
- c. Quais as suposições feitas para escolher o estimador de variância utilizado?

**Exercício 8.7** Seja a população da tabela abaixo:Tabela 8.7: Valores da variável  $y$  para a população

i	I	II	III	IV	V
1	30	63	29	76	54
2	95	24	24	76	55
3	53	39	62	53	73
4	63	38	25	52	37
5	58	79	99	55	96
6	29	36	84	52	79
Total	328	279	323	364	394

Caso não fossem conhecidos os valores da variável representada na tabela, qual seria a melhor alternativa para estimar o total populacional:

- Uma amostra sistemática com  $k = 5$ ?
- Uma AAS de tamanho  $n = 6$ ?
- Uma amostra de Bernoulli com  $p = 0,2$ ? (ESSE PLANO AMOSTRAL AINDA NÃO FOI APRESENTADO!)

Em cada caso calcule a variância do estimador do total populacional para sustentar sua resposta.

**Exercício 8.8** Numa determinada região há 3 comunidades consistindo, respectivamente, de pessoas de descendência Anglo-saxã, Polonesa e Italiana. Está disponível um cadastro atualizado, no qual as pessoas de cada domicílio estão listadas na seguinte ordem: esposo, esposa, filhos por idade e outros residentes. Os domicílios estão listados em ordem em cada rua. O número médio de pessoas por domicílio é 5. Você pode escolher entre uma amostra sistemática com  $K=5$  ou uma AAS com fração amostral de 20%. Para quais das seguintes variáveis você espera, e qual a razão, que a amostragem sistemática seja mais precisa?

- Proporção de pessoas descendentes de Poloneses;
- Proporção de homens;
- Proporção de filhos.

**Exercício 8.9** Um censo foi realizado numa comunidade. A cada 20 domicílios, a partir da seleção aleatória de uma partida entre 1 e 20, foi perguntado o número de anos que a família residia naquele imóvel. Os resultados sumarizados da amostra são dados a seguir:

$$n = 115; \quad N = 2300; \quad \sum_{i=1}^n y_i = 407,1 \quad \text{e} \quad \sum_{i=1}^n y_i^2 = 2011,15$$

a) Descreva o plano amostral utilizado e seus parâmetros; b) Estime o número médio de anos de moradia das famílias nos domicílios; c) Estime a variância do estimador da média, citando a suposição feita sobre a distribuição da variável de interesse na população.

**Exercício 8.10** Os dirigentes de uma associação profissional querem estimar a proporção de membros que são a favor de algumas propostas de alterações no estatuto da entidade. Uma AS foi feita a partir da lista dos associados, selecionando uma partida aleatória entre 1 e 10 e tomando todos os décimos associados da lista, a partir do número sorteado. Foram selecionados 65 sócios no total de 654 existentes. Os sócios da amostra favoráveis às mudanças foram assinalados com  $y_i = 1$  e os contrários com  $y_i = 0$ . Ao final da pesquisa obteve-se  $\sum_{i=1}^{65} y_i = 48$ .

- Estime a proporção,  $p$ , de sócios favoráveis às mudanças no estatuto;
- Estime a margem de erro, com 95% de confiança, para a estimativa de  $p$ .