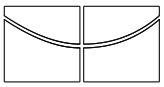


**Universidade de Brasília**

**Instituto de Ciências Exatas**

**Departamento de Estatística**


**Dissertação de Mestrado**

# Implementação Computacional da Amostragem Espacial Adaptável

por


**Iracema Veiga Madeira Mauriz**

**Orientador: Prof. Dr. Alan Ricardo da Silva**



**Novembro de 2013**

**Iracema Veiga Madeira Mauriz**

# Implementação Computacional da Amostragem Espacial Adaptável

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidadede de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

**Universidade de Brasília**

**Brasília, Novembro de 2013.**

# Agradecimentos

Agradeço à Deus por me amparar em todos os momentos, dar-me força interior para superar todas as dificuldades, mostrar-me os caminhos nas horas das decisões mais importantes da minha vida e suprir-me em todas as minhas necessidades. Obrigada por ter colocado pessoas tão especiais em minha vida, sem as quais jamais teria dado conta!

Agradeço a meus pais, Francisco Madeira e Maria José, por sempre acreditarem em minha capacidade, fortalecer-me a fazer o melhor de mim sempre independentemente das circunstâncias, por terem me ensinado tudo, principalmente o amor aos estudos. Obrigada pelo amor incondicional!

Agradeço a meu irmão Tiago, a minha cunhada Patrícia, pois, eles a seus modos sempre se orgulharam de mim e confiaram em meu trabalho. À minha sobrinha, Lara, que desde tão cedo foi uma luz em minha vida. Obrigada pela confiança, pelo incentivo, apoio e carinho!

Agradeço à minha família (tios e tias, primos e primas) a qual amo muito. Obrigada pelo carinho, pela paciência e compreensão!

Agradeço a meu orientador, Prof. Dr. Alan Ricardo da Silva, por acreditar em mim, mostrar-me o caminho da ciência e pesquisa, fazer parte da minha vida nos enriquecedores momentos universitários que envolveram tanto o aprendizado teórico quanto o meu desenvolvimento pessoal, por ser exemplo de profissional que sempre fará parte da minha vida. Obrigada por ter feito do meu sonho o nosso sonho!

Aos Prof. Raul e Prof. George, pelas importantes críticas e sugestões a esta dissertação. Obrigada pelo apoio e incentivo durante o curso com seus exemplos de professores dedicados!

Às minhas amigas de sempre, Laura, Maria Carolina, Marina, Patrícia e Valessa,

por só quererem o meu bem e me valorizarem tanto como pessoa. Obrigada pela amizade!

Agradeço aos amigos que fizeram parte de minha vida sempre me ajudando, aconselhando e incentivando, principalmente a todos os colegas e professores da pós-graduação em estatística. Obrigada pelo convívio, companheirismo e aprendizado!

Finalmente, gostaria de agradecer ao Departamento de Estatística da UnB por abrir as portas para que eu pudesse realizar este sonho que era a minha dissertação de mestrado. Proporcionaram-me mais que a busca de conhecimento técnico e científico, mas uma verdadeira lição de vida!

Obrigada a todos por terem feito esse meu sonho se tornar realidade, afinal ninguém vence sozinho!

# Sumário

# Lista de Figuras

# Lista de Tabelas

# Resumo

Este trabalho teve por objetivo implementar computacionalmente a amostragem espacial adaptável no *software* SAS, uma vez que os seus usuários ainda não tinham uma ferramenta computacional para utilizá-la. Além disso, essa é uma nova técnica que pode ter várias aplicações. A amostragem adaptável é um dos procedimentos que vem sendo estudado e testado em levantamentos de populações raras, como a localização dos minérios, e que exibem padrão de distribuição espacial agregado onde a seleção de unidades amostrais dependem de observações feitas durante a pesquisa. Para isso foi feito um estudo sobre a teoria de amostragem clássica e adaptável, bem como um levantamento de alguns algoritmos de amostragem adaptável já existentes. Assim, definiu-se analisar o caso apresentado por Thompson (1990) onde a distribuição espacial agregada dos dados influencia a seleção amostral dos dados em conglomerados. Dessa forma, desenvolveu-se o algoritmo computacional para a amostragem espacial adaptável por conglomerados e a amostragem espacial adaptável estratificada por conglomerados, obtendo os mesmos resultados para ambos exemplos do autor. Além disso, fez-se outras contribuições para o tema como a análise dos estimadores para a média e para o total com o aumento da população ($N$ grades regulares), bem como a análise de outra forma de seleção amostral, a seleção *QUEEN*, para o caso da amostragem espacial adaptável por conglomerado.

**Palavras Chave**: *amostragem, amostragem adaptável, SAS.*

# Abstract

The objective of the this work is to implement computationally the adaptive spatial sampling on the *software* SAS, since the users do not have a computational tool to use it. Also, this is a new technique that can have multiple applications. Adaptive Sampling is one of the procedures that have been studied and tested in surveys of rare species populations, such as the location of ores, that exhibit spatial pattern household where the selection of sampling units rely on the observations made during the research. For that was made a study on classical and adaptive sampling theory as well as a survey of some existing algorithms of adaptive sampling. Thus, it was decided to analyze the case presented by Thompson (1990) where the clustered distribution of the data influences the selection of sampling data into clusters. Furthermore, was developed a computational algorithm to adaptive spatial cluster sampling and stratified adaptive spatial cluster sampling, getting the same results for both examples of the author. Moreover, it was made further contributions to the subject and the analysis of estimators for the mean and the total with the increase of population ($N$ regular grids), as well as the analysis of other sample selection, such as QUEEN 's selection, to the case of adaptive spatial cluster sampling.

**key words**: *sampling, adaptive sampling, SAS.*

# Introdução

A amostragem tem como objetivo principal obter informações baseando-se no resultado de uma amostra. Segundo Cochran (1977), a teoria de amostragem foi desenvolvida a fim de tornar a amostragem mais eficiente, isto é, produzir estimativas mais precisas com menor custo possível. Assim, o problema básico de qualquer procedimento de amostragem é a obtenção de estimativas fidedignas de alguma característica da população de interesse, tomando como base somente parte dessa população.

Um dos procedimentos que vem sendo estudado e testado em levantamentos de populações de espécies raras e que exibem padrão de distribuição espacial agregado é a amostragem adaptável. Nesse tipo de amostragem, a seleção de unidades amostrais depende de observações feitas durante a pesquisa; pois se for satisfeito um critério, então a amostra vizinha é adicionada à amostra inicial. Assim, essa amostragem tem vantagens como um maior aproveitamento da amostra, maior intensidade da amostragem dependendo das observações feitas durante a pesquisa, além de poder ajudar a encontrar os máximos locais (Thompson e Seber, 1996).

Na literatura da amostragem adaptável, verifica-se alguns exemplos de algoritmos como: a análise dos efeitos das mutações nas propriedades de dobragem do *RNA* para decifrar os princípios de condução e a evolução molecular para conceber novas moléculas, ou seja, um algoritmo de amostragem adaptável imparcial que permite *RNAmutants* para as regiões da amostra da paisagem mutacional mal coberto por técnicas anteriores (Waldispühl e Ponty, 2011); para os desenhos de estudo clínico centrado em projetos de adaptação de dois estágios com o tamanho de amostra re-estimativa (Chang, 2008; Robinson, 2009); em aplicações de mineração com grande quantidade de dados, onde a amostragem aleatória é difícil de usar devido à difi-

culdade de determinar um tamanho de amostra apropriado (Domingo et al., 2002); ou quando os algoritmos para minar toda a informação disponível em um grande banco de dados é proibitivo devido a restrições computacionais (tempo e memória) (Satyanarayana e Davidson).

No caso apresentado por (Thompson, 1990) foi considerado uma outra forma de amostragem adaptável, onde a distribuição espacial agregada dos dados influencia a seleção amostral dos dados em conglomerados. Assim, esse desenho amostral em que o procedimento de seleção de unidades pode ser adicionado à amostra inicial, observando uma área e a sua distribuição espacial, será denominado nesse trabalho de amostragem espacial adaptável.

Um exemplo simples para esse tipo de amostragem é dado pela Figura 1. Inicialmente pretende-se analisar uma determinada área, Figura 1 (a). Depois, é desenhada uma grade regular em cima da área a ser pesquisada, Figura 1 (b). Em seguida, são selecionadas $n$ unidades (quadrados) pelo método AAS. Nesse exemplo, verifica-se pela Figura 1 (c) que a amostra inicial é formada por 5 unidades (total de quadrados na grade em vermelho), que foram selecionadas por meio de uma $AAS$ de um total de $N = 121$ unidades (que representam o número total de quadrados da grade regular onde cada lado tem tamanho igual a 11, ou seja, $11 \times 11 = 121$), com total de 10 unidades de interesse.

Selecionando os vizinhos das unidades iniciais que contém pelo menos uma unidade na amostra inicial, Figura 1 (c), obtem-se a amostra final, Figura 1(d). A unidade superior tem um elemento, que intercepta a rede com $m_1 = 2$ unidades, que contém um total de $y_1^* = 5$ unidades de interesse (1+4=5), ou seja, os dois quadrados selecionados são os de número 47 e 48 com 1 e 4 unidades de interesse, respectivamente. Outro ponto em que se verifica a unidade dentro do polígono que intercepta a rede com $m_2 = 1$ unidades, contém $y_2^* = 5$ unidades, ou seja, o quadrado 25 tem as 5 unidades de interesse. Para as outras 3 unidades da amostra inicial, o valor de $y_i = 0$ e $m_i = 1$, representando os quadrados 5, 86 e 109 que não têm elementos de interesse.

Dessa forma, a amostragem espacial adaptável oferece uma solução viável para o problema de longa data de estimar a abundância de populações raras e está rapi-

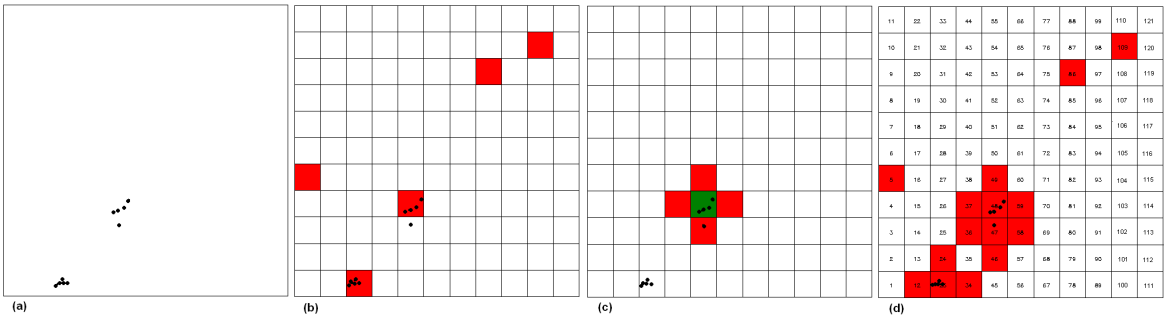


Figura 1: Exemplo da Amostragem Espacial Adaptável

damente ganhando destaques no campo das ciências naturais e sociais (Seber, 1986; Ramsey e Seber, 1992; Brown, 1994, 1996; Khan e Muttlak, 2002; Stein e Ettema, 2003; Sengupta e Sengupta, 2011; Jain e Chang, 2004; Thompson, 2011; yu). O problema é que os procedimentos de adaptação são mais complicados para projetar e analisar, assim como as implementações computacionais são poucas devido a complexidade dos algoritmos de análise espacial (Thompson, 2011).

Essa implementação requer pelo menos três passos: o desenvolvimento computacional do desenho de grade regular (ou do inglês *grid*); seleção de áreas específicas da grade, isto é, identificar uma amostra e verificar em qual parte dessa grade estão os dados; identificar os vizinhos das áreas selecionadas: superiores, inferiores, da direita e da esquerda. Sendo assim, o intuito desta dissertação é implementar computacionalmente a amostragem espacial adaptável no *software* SAS.

No Capítulo 1 serão feitas algumas considerações iniciais sobre a teoria de amostragem, como os estimadores não viesados para cada amostragem: amostragem aleatória simples, estratificada, por conglomerados. No Capítulo 2 será abordada a amostragem espacial adaptável como suas vantagens e definições, sendo abordado a amostragem espacial adaptável por conglomerados e amostragem espacial adaptável estratificada por conglomerado. No Capítulo 3 será apresentado o algoritmo desenvolvido com a especificação do desenho da grade regular, a seleção das áreas específicas da grade, a identificação dos vizinhos das áreas selecionadas, bem como algumas contribuições como o uso de matriz *queen* e *rook*. No Capítulo 4 serão apresentados os resultados do algoritmo utilizando bases de dados simuladas, além da comparação entre a variação dos tamanhos populacionais, $N$, das grades. Por fim,

no Capítulo 5 serão apresentadas as conclusões do trabalho, algumas recomendações para trabalhos futuros e as limitações do trabalho.

# Capítulo 1

# Teoria de Amostragem

## 1.1 Introdução

A base conceitual da amostragem foi desenvolvida durante o século XX para estudar qualquer conjunto de indivíduos. O método era utilizado para realizar a enumeração desses e os estudos estatísticos tinham como objetivo inventariar os recursos das nações para fins militares e tributários e sua teoria foi desenvolvida desde então (Hansen e Hurwitz, 1943; Cornfield, 1944; Tukey, 1950; Kish, 1965; Raj, 1968; Cochran, 1977; Foreman, 1991; Lohr, 1999; Thompson, 2002).

A noção inicial que uma pessoa tem sobre amostragem é fato que decorre de suas experiências cotidianas como, por exemplo, quando um indivíduo ao testar a temperatura de seu prato de sopa experimenta apenas uma parte dessa. Assim, a amostragem tem o objetivo principal de obter informações sobre o todo, baseando-se no resultado de uma amostra, bem como fazer o uso de amostras que produzam resultados amostrais confiáveis e livres de viéses (Bolfarine e Bussab, 2005).

Além disso, a amostragem é uma técnica que possui vantagens como a redução de custo ao ser determinado o tamanho amostral; a maior velocidade ao se analisar parte dos dados e não a sua totalidade; assim como maior amplitude e exatidão ao se utilizar pessoas especializadas, proporcionando resultados um pouco mais precisos (Cochran, 1977). Dessa forma, o papel da teoria da amostragem é fazer uma amostragem mais eficiente com o menor custo possível, acarretando estimativas mais precisas.

Pode-se dizer ainda que dois aspectos afetam a quantidade de informação contida em uma amostra e, consequentemente, a precisão da tomada de decisão nos procedimentos de inferência. O primeiro deles é o tamanho da amostra selecionada a partir da população. O segundo é a quantidade de variação dos dados, a qual pode ser frequentemente controlada por meio do método de seleção amostral (Scheaffer et al., 1996).

Os tipos de amostragem a serem analisados nesse Capítulo serão os da amostragem probabilística clássica: amostragem aleatória simples, amostragem estratificada, amostragem por conglomerados em um ou dois estágios. Inicia-se, então, esse trabalho com a introdução do conceito de função indicadora.

## 1.2 Função Indicadora

A função indicadora $I$ pode asumir os valores 1, se a unidade de interesse estiver incluída na amostra, e 0 caso contrário. Essa função em amostragem é muito importante, pois em geral o interesse da amostragem é saber se a unidade está ou não incluída na amostra. Outra vantagem de usar a função indicadora é a sua aplicação a qualquer evento, bem como a facilidade matemática nos cálculos de suas derivações das médias e variâncias dos estimadores (Thompson e Seber, 1996).

Assim, $I_A(\omega) = 1$ se e somente se $\omega \ \epsilon \ A$, isto é, $P(I_A(\omega) = 1) = P(A)$ e $P(I_A(\omega) = 0) = P(\overline{A})$. Com isso, pode ser obtido a sua esperança, $E[I_A(\omega)] = 1 \times P(A) + 0 \times P(\overline{A})$, a sua variância

$$Var[I_A(\omega)] = E[I_A^2] - (E[I_A])^2 = E[I_A] - [P(A)]^2 = P(A)[1 - P(A)] \qquad (1.1)$$

e sua covariância,

$$Cov[I_A, I_B] = E[I_A I_B] - E[I_A]E[I_B] = E[I_{A\cap B}] - [P(A)][P(B)] \qquad (1.2)$$

Uma das aplicações mais comumente relacionadas ao uso da função indicadora é o cálculo da média e da variância de uma variável indicadora como, por exemplo, $Z = \sum_{i=1}^{N} z_i I_i$. Seja $I_i$, a variável aleatória (v.a.), o indicador da variável que tem

o seu valor igual a 1 com probabilidade $\Pi_i$ e $I_iI_j = 1$ com probabilidade $\Pi_{ij}$, onde $i \neq j$ (Thompson e Seber, 1996). Assim,

$$E[Z] = E\left[\sum_{i=1}^{N} z_i I_i\right] = \sum_{i=1}^{N} z_i E[I_i] = \sum_{i=1}^{N} \Pi_i \tag{1.3}$$

A partir de (1.1) e (1.2) obtem-se a variância de $Z$,

$$\begin{aligned}
Var[Z] &= \sum_{i=1}^{N} var[z_i I_i] + \sum_{i=1}^{N}\sum_{i\neq j} cov[z_i I_i z_j I_j] \\
&= \sum_{i=1}^{N} z_i^2 P(A)[1-P(A)] + \sum_{i=1}^{N}\sum_{i\neq j} z_i z_j [P(A\cap B) - P(A)P(B)] \\
&= \sum_{i=1}^{N} z_i^2 \Pi_i(1-\Pi_i) + \sum_{i=1}^{N}\sum_{i\neq j} z_i z_j \Pi_{ij} - \Pi_i\Pi_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} z_i z_j (\Pi_{ij} - \Pi_i\Pi_j)
\end{aligned} \tag{1.4}$$

Dado que $\Pi_{ii} = \Pi_i$, $E[I_i] = E[I_i^2] = \Pi_i$ e $E[I_iI_j] = \Pi_{ij}$, obtem-se o seguinte estimador não-viesado para a Equação (1.4)

$$\begin{aligned}
\widehat{Var}[Z] &= \sum_{i=1}^{n} \frac{z_i^2 \Pi_i(1-\Pi_i) I_i I_j}{E[I_iI_j]} + \frac{\sum_{i=1}^{n}\sum_{i\neq j} z_i z_j (\Pi_{ij} - \Pi_i\Pi_j) I_i I_j}{E[I_iI_j]} \\
&= \sum_{i=1}^{n} \frac{z_i^2 \Pi_i(1-\Pi_i) I_i}{\Pi_i} + \frac{\sum_{i=1}^{n}\sum_{i\neq j} z_i z_j (\Pi_{ij} - \Pi_i\Pi_j) I_i I_j}{\Pi_{ij}} \\
&= \sum_{i=1}^{n} z_i^2 (1-\Pi_i) I_i + \sum_{i=1}^{n}\sum_{i\neq j} z_i z_j \frac{(\Pi_{ij} - \Pi_i\Pi_j)}{\Pi_{ij}} I_i I_j \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} z_i z_j I_i I_j \left(\frac{\Pi_{ij} - \Pi_i\Pi_j}{\Pi_{ij}}\right)
\end{aligned} \tag{1.5}$$

Na próxima Seção apresenta-se algumas inferências sobre a teoria de amostragem aleatória simples com e sem reposição, o que ajudará no desenvolvimento da teoria da amostragem espacial adaptável.

## 1.3 Amostragem Aleatória Simples

A amostragem aleatória simples ($AAS$) é um método mais simples e importante para a seleção de uma amostra que consiste em selecionar $n$ unidades de $N$ tal que qualquer uma das $C_n^N$ das amostras distintas tenham uma chance igual de ser sorteada. O seu plano pode ser descrito a partir de um universo $U = \{1, 2, \ldots, N\}$ em três passos (Bolfarine e Bussab, 2005):

- Utilizando-se um processo aleatório (seja por urna, tabela de números aleatórios, ou outro), sorteia-se com igual probabilidade um elemento da população $N$;
- Repete-se o processo anterior até que sejam sorteadas $n$ unidades;
- Caso seja permitido o sorteio de uma unidade mais de uma vez, tem-se o processo AAS com reposição ($AAS_C$). Caso contrário, tem-se o processo AAS sem reposição ($AAS_S$).

Uma consequência da definição de $AAS$ é que todos os indivíduos dessa população tem a mesma chance de ser selecionado; mas, esse fato não pode ser incorporado na definição de $AAS$, pois isso não implica que todas as amostras de tamanho $n$ tenham a mesma chance de ser selecionada (Scheaffer et al., 1996). Outros dois fatos importantes desse desenho amostral é que toda a amostra de unidades a serem observadas é fixada antes de se iniciar a pesquisa (Cochran, 1977) e que a probabilidade de um elemento da população estar na amostra é conhecido.

A seguir serão consideradas algumas definições. Seja $N$, o tamanho da população que varia de $(1, 2, \ldots, N)$; $n$, o tamanho da amostra; $E = \{E_1, \ldots, E_N\}$, os elementos da população; $A = \{e_1, \ldots, e_n\}$, os elementos da amostra; $Y_i$, a variável de interesse ou medida em cada elemento $E_j$; $P = \{Y_1, Y_2, \ldots, Y_N\}$, a variável de interesse da população e $A = \{y_1, y_2, \ldots, y_n\}$, a variável de interesse da amostra.

Sejam os parâmetros de interesse $\tau = \sum_{i=1}^{N} Y_i = N\mu$, o total de números da unidade de interesse; $\mu = \overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$, a média por unidade; $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \mu)^2$, a variância populacional (caso com reposição); $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \mu)^2$, a variância populacional (caso sem reposição).

Sejam os estimadores dos parâmetros de interesse dados por $\widehat{\tau} = N\overline{y}$, o estimador do total; $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, o estimador da média populacional; $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$, o estimador das variâncias. Esses parâmetros e estimadores serão usados na próxima Seção no estudo da seleção com reposição e na Seção seguinte no estudo da seleção sem reposição.

### 1.3.1 Amostragem Aleatória Simples com Reposição

A $AAS_C$ é realizada a partir de um universo $U = \{1, 2, \ldots, N\}$ em três passos (Bolfarine e Bussab, 2005):

- Utilizando-se um processo aleatório (seja por urna, tabela de números aleatórios, ou outro), sorteia-se com igual probabilidade uma das $N$ unidades da população;
- Repõe-se essa unidade na população e sorteia-se um elemento seguinte;
- Repete-se o processo até que $n$ unidades tenham sido sorteadas.

Para esse plano amostral é fácil verificar que a variável $f_i$, número de vezes que a unidade $i$ aparece na amostra, segue uma distribuição Binominal com parâmetros $n$ e $p = \frac{1}{N}$. Assim, tem-se a esperança e variância dessa variável dada por:

$$E[f_i] = np = \frac{n}{N} \tag{1.6}$$

$$Var[f_i] = np(1-p) = \frac{n}{N}\left(1 - \frac{1}{N}\right) \tag{1.7}$$

No plano de $AAS_C$ cada tentativa é independente e qualquer um dos $N$ elementos populacionais tem a mesma probabilidade $\frac{1}{N}$ de ser sorteado, caracterizando para as variáveis, $(f_1, f_2, \ldots, f_N)$, uma distribuição multinomial com parâmetros $\left(n; \frac{1}{N}, \ldots, \frac{1}{N}\right)$. Assim, sabendo-se a distribuição da variável e de (1.2) pode-se obter a covariância dessas variáveis.

$$cov(I_i(l), I_j(l)) = E(I_i(l) \times I_j(l)) - E(I_i(l)) \times E(I_j(l)) = 0 - E(I_i(l)) \times E(I_j(l)) = -P_i \times P_j \tag{1.8}$$

onde $E(I_i(l) \times I_j(l)) = 0$, pois não se pode retirar o elemento $i$ e $j$ simultaneamente.

Dado que $P_i = P_j = \frac{1}{N}$ e de (1.8) obtem-se

$$cov(f_i, f_j) = -nP_i \times P_j = -\frac{n}{N^2} \tag{1.9}$$

Seja $t = \sum_{j=1}^{n} Y_j = \sum_{j=1}^{N} f_j Y_j$, onde $f_j$ é uma variável aleatória que indica quantas vezes o elemento $i$ aparece na amostra. Então, obtem-se algumas propriedades dessa estatística como a sua esperança e variância.

$$E[t] = E\left[\sum_{j=1}^{N} f_j Y_j\right] = \sum_{j=1}^{N} Y_j E(f_j) = E(f) \sum_{j=1}^{N} Y_j = \frac{n}{N} \sum_{j=1}^{N} Y_j = n\mu \tag{1.10}$$

Dado que $var(f) = \frac{n}{N}\left(1 - \frac{1}{N}\right) = \frac{n(N-1)}{N^2}$, $(N\mu)^2 = (\sum_{j=1}^{N} Y_j)^2 = \sum_{j=1}^{N} Y_j^2 + \sum \sum_{i \neq j} Y_i Y_j$ e $(Y_1 + Y_2)^2 = Y_1^2 + Y_1 Y_2 + Y_2 Y_1 + Y_2^2$, então a variância de $t$ é dada por:

$$\begin{aligned}
var[t] &= var\left(\sum_{j=1}^{N} f_j Y_j\right) = \sum_{j=1}^{N} Y_j^2 var(f_j) + \sum \sum_{i \neq j} Y_i Y_j cov(f_i, f_j) \\
&= var(f) \sum_{j=1}^{N} Y_j^2 + cov(f, f') \sum \sum_{i \neq j} Y_i Y_j \\
&= var(f) \sum_{j=1}^{N} Y_j^2 + cov(f, f')(N^2\mu^2 - \sum_{j=1}^{N} Y_j^2) \\
&= \frac{n(N-1)}{N^2} \sum_{j=1}^{N} Y_j^2 - \frac{n}{N^2}(N^2\mu^2 - \sum_{j=1}^{N} Y_j^2) \\
&= \frac{n}{N^2}\left((N-1)\sum_{j=1}^{N} Y_j^2 - N^2\mu^2 + \sum_{j=1}^{N} Y_j^2)\right) = \frac{n}{N^2} N \left(\sum_{j=1}^{N} Y_j^2 - N\mu\right) \\
&= n\sigma^2
\end{aligned} \tag{1.11}$$

Com isso, pode-se obter estimadores não-viesados para a média populacional, $\mu$.

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} Y_i = \frac{t}{n} = \widehat{\mu} \tag{1.12}$$

Verificando que $\overline{y}$ é um estimador não-viesado da média populacional $\mu$ dentro do plano $AAS_C$, tem-se que

$$E[\overline{y}] = E\left[\frac{t}{n}\right] = \frac{1}{n}E[t] = \frac{1}{n}n\mu = \mu \tag{1.13}$$

A variância desse estimador é dada por

$$var[\overline{y}] = var\left[\frac{t}{n}\right] = \frac{1}{n^2}var(t) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \tag{1.14}$$

Similarmente, pode ser encontrado um estimador não-viesado para o total populacional, $\tau$ e a sua variância.

$$T = N\overline{y} = \frac{N\sum_{i=1}^{n} y_i}{n} = \frac{N}{n}t = \widehat{\tau} \tag{1.15}$$

$$var[T] = var[\widehat{\tau}] = var\left(\frac{Nt}{n}t\right) = \frac{N^2}{n^2}var(t) = \frac{N^2}{n^2}n\sigma^2 = \frac{N^2\sigma^2}{n} \tag{1.16}$$

Dentro desse plano amostral, a estatística $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{y})^2$ é um estimador não-viesado para a variância populacional $\sigma^2$, ou seja,

$$\begin{aligned}(n-1)s^2 &= \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(y_i^2 - 2Y_i\overline{y} + \overline{y}^2) = \sum_{i=1}^{n} y_i^2 - 2n\overline{y}\overline{y} + n\overline{y}^2 \\ &= \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 = \sum_{i=1}^{n} y_i^2 - \frac{nt^2}{n^2} = \sum_{i=1}^{n} y_i^2 - \frac{nt^2}{n} = \sum_{i=1}^{n} y_i^2 - \frac{t^2}{n}\end{aligned} \tag{1.17}$$

Dado que $E\left(\sum_{i=1}^{n} y_i^2\right) = E[\sum_{i=1}^{N} f_i Y_i] = E(f)\sum_{i=1}^{N} Y_i^2 = \frac{n}{N}\sum_{i=1}^{N} Y_i^2 = \frac{n(N(\sigma^2+\mu^2))}{N} = n(\sigma^2+\mu^2)$, que $\sigma^2 = \frac{\sum_{i=1}^{N}(Y_i-\overline{y})^2}{N} \Rightarrow N\sigma^2 = \sum_{i=1}^{N} Y_i^2 - N\overline{Y}^2 \Rightarrow N\sigma^2 + N\mu^2 = \sum_{i=1}^{N} Y_i^2$ e que $E[t^2] = var[t] + [E(t)]^2 = n\sigma^2 + n^2\mu^2 = n(\sigma^2 + n\mu^2)$, obtem-se

$$\begin{aligned}E[(n-1)s^2] &= E\left[\sum_{i=1}^{n} y_i^2 - \frac{t^2}{n}\right] = E\left[\sum_{i=1}^{n} y_i^2\right] - E\left[\frac{t^2}{n}\right] \\ &= E\left[\sum_{i=1}^{n} y_i^2\right] - \frac{1}{n}E[t^2] = n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 \\ &= \sigma^2(n-1)\end{aligned} \tag{1.18}$$

O estimador não-viesado da variância da média e do total da população, $var[\overline{y}]$ e de de $var[T]$, respectivamente, é dado substituindo $\sigma^2$ por $s^2$, ou seja,

$$\widehat{var}(\overline{y}) = \frac{s^2}{n} \tag{1.19}$$

$$\widehat{var}(T) = N^2\frac{s^2}{n} \tag{1.20}$$

Pelo teorema central do limite (TCL), sabe-se que à medida que o tamanho da amostra aumenta, as distribuições de $\overline{y}$ e de $\widehat{\tau}$ se aproximam da distribuição normal. Assim, para $n$ suficientemente grande, tem-se que

$$\frac{\overline{y} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1) \tag{1.21}$$

$$\frac{\widehat{\tau} - \tau}{N\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1) \tag{1.22}$$

Consequentemente, os intervalos de confiança aproximados para $\overline{y}\ \epsilon\ \tau$, ou seja,

$$P\left(\overline{y} - z_{\alpha/2}\sqrt{\frac{s^2}{n}} \leq \mu \leq \overline{y} + z_{\alpha/2}\sqrt{\frac{s^2}{n}}\right) \simeq 1 - \alpha \tag{1.23}$$

de onde segue que o intervalo de confiança (IC) para $\mu$ com coeficiente de confiança aproximadamente igual a $1-\alpha$ é dado por $\left(\overline{y} - z_{\alpha/2}\sqrt{\frac{s^2}{n}}; \overline{y} + z_{\alpha/2}\sqrt{\frac{s^2}{n}}\right)$. Isso significa que de 100 amostras $AAS$ e de 100 IC baseados nessas amostras, aproximadamente $100(1-\alpha)\%$ dos intervalos devem conter $\mu$.

Por fim, para determinar o tamanho da amostra $n$, de tal forma que o estimador obtido tenha um erro máximo de estimação igual a $\varepsilon$, com determinado grau de confiança, tem-se que $P(|\overline{y} - \mu| \leq \varepsilon) = P\left(|\overline{y} - \mu| \leq z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right) \approx 1 - \alpha$, isto é, $n = \frac{z^2_{\alpha/2}\sigma^2}{\varepsilon^2}$.

### 1.3.2 Amostragem Aleatória Simples sem Reposição

A $AAS_S$ é realizada a partir de um universo $U = \{1, 2, \ldots, N\}$ em três passos (Bolfarine e Bussab, 2005):

- Utilizando-se um processo aleatório (seja por urna, tabela de números aleatórios, ou outro), sorteia-se com igual probabilidade uma das $N$ unidades da população;
- Sorteia-se um elemento seguinte, com o elemento anterior sendo retirado da população;
- Repete-se o processo até que $n$ unidades tenham sido sorteadas.

Sendo assim, a $AAS_S$ funciona de modo idêntico à $AAS_C$, a não ser pela não recolocação do elemento sorteado. Dessa maneira, o elemento populacional que for selecionado para fazer parte da amostra será único. Portanto, a variável $f_i$, número de vezes que a unidade $i$ aparece na amostra segue a distribuição Bernoulli.

Com isso, todas as amostras de tamanho $n$ são igualmente prováveis de acontecer. Assim, qualquer uma dessas amostras têm a probabilidade $\frac{1}{C_n^N}$ de ser selecionada. Como o número de amostras contendo a unidade $i$ é $C_{n-1}^{N-1}$, então a probabilidade de $I_i = 1$ resulta na fração amostral, $\frac{n}{N}$, ou seja,

$$P(I_i = 1) = \frac{C_{n-1}^{N-1}}{C_n^N} = \frac{\frac{(N-1)!}{(n-1)!(N-1-n+1)!}}{\frac{N!}{n!(N-n)!}} = \frac{(N-1)!n(n-1)!(N-n)!}{(n-1)!N(N-1)!(N-n)!} = \frac{n}{N} = \Pi_i \tag{1.24}$$

Consequentemente, obtem-se a esperança e variância da variável $f_i$, bem como as suas probabilidades de inclusão e covariância.

$$E(f_i) = 1 \times P(f_i = 1) + 0 \times P(f_i = 0) = P(f_i = 1) = \frac{n}{N} \tag{1.25}$$

$$Var(f_i) = p(1-p) = \frac{n}{N} \times (1 - \frac{n}{N}) = \frac{n}{N}\left(\frac{N-n}{N}\right) \tag{1.26}$$

$$\Pi_i = P(f_i = 1) = \frac{n}{N} \tag{1.27}$$

$$\Pi_{ij} = P(f_i = 1 \cap f_i = 0) = \frac{n}{N} \times \frac{n-1}{N-1} \tag{1.28}$$

$$
\begin{aligned}
cov(f_i, f_j) &= E(f_i f_j) - E(f_i)E(f_j) = \frac{n}{N} \times \frac{n-1}{N-1} - \frac{n}{N} \times \frac{n}{N} = \frac{n}{N} \times \frac{n-1}{N-1} - \frac{n^2}{N^2} \\
&= \frac{N^2(n^2-n) - n^2(N^2-N)}{N^2(N^2-N)} = \frac{-N^2 n + Nn^2}{N^2(N^2-N)} = \frac{-Nn(n-N)}{N^2(N^2-N)} \\
&= \frac{-n}{N^2} \times \frac{(N-n)}{(N-1)} \qquad (1.29)
\end{aligned}
$$

onde $\frac{(N-n)}{(N-1)} \approx 1$ quando $N$ tende a ser grande ao ser comparado com o tamanho da amostra, $n$. Essa fração é denominada de correlação para populações finitas (ou do inglês *fpc - finite population correction*).

Analogamente à $AAS_C$, obtem-se a esperança e variância da estatística $t$ para o caso da $AAS_S$

$$
E[t] = E\left[\sum_{i=1}^{N} f_i Y_i\right] = E(f)\sum_{i=1}^{N} Y_i = \frac{n}{N} \times \sum_{i=1}^{N} Y_i = n\mu \qquad (1.30)
$$

$$
\begin{aligned}
var(t) &= var\left(\sum_{i=1}^{N} f_i Y_i\right) = var(f_i)\sum_{i=1}^{N} Y_i^2 + cov(f_i, f_j)\sum\sum_{i\neq j} Y_i Y_j \\
&= \frac{n}{N}\left(1 - \frac{n}{N}\right)\sum_{i=1}^{N} Y_i^2 - \frac{n}{N^2}\left(\frac{N-n}{N-1}\right)\left(N^2\mu^2 - \sum_{i=1}^{N} Y_i^2\right) \\
&= \frac{n}{N}\left(\frac{N-n}{N}\right)\sum_{i=1}^{N} Y_i^2 - \frac{n}{N^2}\left(\frac{N-n}{N-1}\right)(N^2\mu^2 + \frac{n}{N^2}\left(\frac{N-n}{N-1}\right)\sum_{i=1}^{N} Y_i^2 \\
&= \frac{n(N-n)}{N^2}\left(\sum_{i=1}^{N} Y_i^2 - \frac{N^2\mu^2}{N-1} + \frac{\sum_{i=1}^{N} Y_i^2}{N-1}\right) \\
&= \frac{n(N-n)}{N^2}\left(\frac{N\sum_{i=1}^{N} Y_i^2 - \sum_{i=1}^{2} Y_i^2 - N^2\mu^2 + \sum_{i=1}^{N} Y_i^2}{N-1}\right) \\
&= \frac{n(N-n)}{N^2}\left(\frac{N\sum_{i=1}^{N} Y_i^2 - N^2\mu^2}{N-1}\right) = \frac{n(N-n)}{N^2} N\left(\frac{\sum_{i=1}^{N} Y_i^2 - N\mu^2}{N-1}\right) \\
&= \frac{n(N-n)}{N} S^2 = n\left(1 - \frac{n}{N}\right) S^2 = n(1-f)S^2 \qquad (1.31)
\end{aligned}
$$

onde $f = \frac{n}{N}$. Para um valor de $N$ grande, $lim_{N\to\infty} \frac{n}{N} \to 0$ e $n(1-0)S^2 = nS^2$ e $S^2 = \frac{\sum_{i=1}^{N} Y_i^2 - N\mu^2}{N-1} \approx \frac{\sum_{i=1}^{N} Y_i^2 - N\mu^2}{N} = \sigma^2$, então $var(t) = n\sigma^2$. Isso significa que se $N$ é grande, as $AAS_C$ e $AAS_S$ são equivalentes matematicamente.

Calculando a esperança e a variância dessa unidade obtida pela $AAS_S$, tem-se que $\overline{y}$ e $s^2$ é não-viesado para $\mu$ e para $S^2$, respectivamente, como é provado a seguir.

$$E[\overline{y}] = \frac{1}{n}\sum_{i=1}^{N} y_i P(I_i = 1) = \frac{1}{n}\sum_{i=1}^{N} y_i \frac{n}{N} = \frac{1}{N}\sum_{i=1}^{N} y_i = \mu \tag{1.32}$$

$$var[\overline{y}] = var\left(\frac{t}{n}\right) = \frac{1}{n^2} var(t) = \frac{n(1-f)S^2}{n^2} = \frac{(1-f)S^2}{n} = \frac{N^2(1-f)S^2}{n} \tag{1.33}$$

Dessa forma, $\widehat{\mu} = \overline{y}$, ou seja, $\overline{y}$ é não-viesado tanto para $AAS_S$ quanto para a $AAS_C$ (Cochran, 1977).

Similarmente, pode ser encontrado um estimador não-viesado para o total populacional, $\tau$ e a sua variância

$$T = N\overline{y} = \frac{Nt}{n} = \widehat{\tau} \tag{1.34}$$

$$var(T) = var(\widehat{\tau}) = N^2 var\left(\frac{t}{n}\right) = \frac{N^2(1-f)S^2}{n} \tag{1.35}$$

Analogamente, o estimador não-viesado da variância da média e do total da população, $var[y]$ e $var[T]$, respectivamente é dado por:

$$\widehat{var}(\overline{y}) = \frac{(1-f)s^2}{n} \tag{1.36}$$

$$\widehat{var}(T) = \widehat{var}(\widehat{\tau}) = \frac{N^2(1-f)s^2}{n} \tag{1.37}$$

Como no caso $AAS_C$, pelo TCL sabe-se que para $n$ suficientemente grande, tem-se que

$$\frac{\overline{y} - \mu}{\sqrt{\frac{(1-f)\sigma^2}{n}}} \sim N(0,1) \tag{1.38}$$

$$\frac{\widehat{\tau} - \tau}{N\sqrt{\frac{(1-f)\sigma^2}{n}}} \sim N(0,1) \tag{1.39}$$

e os intervalos de confiança aproximados para $\overline{y}\ \epsilon\ \tau$, ou seja,

$$P\left(\overline{y} - z_{\alpha/2}\sqrt{(1-f)\frac{s^2}{n}} \leq \mu \leq \overline{y} + z_{\alpha/2}\sqrt{(1-f)\frac{s^2}{n}}\right) \simeq 1-\alpha \tag{1.40}$$

onde para $n < 50$, utiliza-se a tabela $t$ de*Student* com $n-1$ graus de liberdade.

Quando se deseja determinar o tamanho da amostra $n$, dispõem-se de $var(\overline{y}) = (1-f)\frac{s^2}{n} = \frac{s^2}{n'}$, onde $n' = \frac{n}{(1-f)}$ e $\epsilon = z_{\alpha/2}\sqrt{\frac{s^2}{n}}$. Assim, $n' = \frac{z^2_{\alpha/2}\sigma^2}{\varepsilon^2} = \frac{n}{(1-f)} \Rightarrow n = \frac{n'}{1+\frac{n'}{N}}$.

### 1.3.3 Comparação entre $AAS_C$ e $AAS_S$

O EPA - Efeito do Planejamento Amostral (ou do inglês *Deff - Design Effect*), é uma importante ferramenta de comparação entre dois tipos de amostragem para verificar qual plano é o melhor. Esse cálculo é feito por meio da comparação das variâncias de um plano considerado padrão. No caso da $AAS_C$ e da $AAS_S$, sabe-se que a estatística $\overline{y}$ é, em ambos os casos, um estimador não-viesado para a média $\mu$. Com isso, o *Deff* para esses dois tipos de amostragem é dado por (Cochran, 1977):

$$\begin{aligned} Deff &= \frac{var(AAS_S(\overline{y}))}{var(AAS_C(\overline{y}))} = \frac{(1-f)\frac{S^2}{n}}{\frac{\sigma^2}{n}} = \frac{\left(\frac{N-n}{N}\right) \times \frac{S^2}{n}}{\frac{\sigma^2}{n}} \\ &= \frac{\left(\frac{N-n}{N}\right) \times \frac{\sum_{i=1}^{N}(y_i-\overline{y})^2}{N-1} \times \frac{1}{n}}{\frac{\sigma^2}{n}} = \frac{\left(\frac{N-n}{N-1}\right) \times \frac{\sigma^2}{n}}{\frac{\sigma^2}{n}} = \frac{N-n}{N-1} \end{aligned} \tag{1.41}$$

onde *Deff*$< 1$, para $n > 1$. Isso significa que a $AAS_S$ é sempre melhor do que a $AAS_C$, pois quando o *Deff*$< 1$, o plano amostral do numerador é mais eficiente que o padrão. Assim, se confirma a teoria ao verificar que, intuitivamente, uma amostra sem reposição representa melhor uma população do que uma com reposiçcão.

Além disso, o plano de $AAS_S$ é muito mais interessante na prática, pois não se pode ganhar mais informações se uma mesma unidade aparecer mais de uma vez na amostra. Mas a $AAS_C$ tem vantagens como a independência entre as unidades sorteadas, o que facilita os cálculos das propriedades populacionais de interesse.

Na Seção seguinte será abordada a teoria de amostragem estratificada que será de suma importância no desenvolvimento da amostragem espacial adaptável estra-

tificada por conglomerado.

## 1.4 Amostragem Estratificada

Nessa Seção, objetiva-se abordar a amostragem estratificada clássica ($AE$) onde a população de $N$ unidades é primeiro dividida em $L$ estratos. Isso significa que as subpopulações são compostas de $N_1, N_2, \ldots, N_L$ unidades que não se sobrepõem, mas juntas resultam na população, isto é, $N_1 + N_2 + \ldots + N_L = N$ e que os tamanhos das amostras dentro de cada estrato são denominados por $n_1, n_2, \ldots, n_L$, ou seja $n_1 + n_2 + \ldots + n_L = n$. Os benefícios dessa amostragem, que é utilizada para resolver problemas como a melhoria das estimativas e produzir estimativas para toda a população e subpopulações, são obtidos se os valores de cada subpopulação, $N_h$, onde $h = 1, \ldots, L$, são conhecidos (Cochran, 1977).

Para esse tipo de amostragem supõem-se que seja possível dividir uma população heterogênea em subpopulações ou estratos internamente homogêneas, desejando obter, uma estimativa mais precisa da média de qualquer estrato com uma amostra menor de cada estrato (Scheaffer et al., 1996). Consequentemente, essas estimativas podem ser combinadas dentro de uma estimativa precisa para toda a população.

A execução de um plano de $AE$ possui quatro passos (Bolfarine e Bussab, 2005): A divisão da população em subpopulações bem definidas, estratos; de cada estrato retira-se uma amostra, geralmente independente; em cada amostra, usam-se estimadores convenientes para os parâmetros do estrato; monta-se para a população um estimador combinando os estimadores para cada estrato, determinando então as suas propriedades.

Seja $h$ o estrato e $i$ a unidade dentro do estrato, então a notação da $AE$ dada por:

- $N_h$: Número total de unidades na população;
- $n_h$: Número total de unidades na amostra;
- $y_{hi}$: Valor obtido para a $i$-ésima unidade do estrato $h$;
- $N = \sum_{h=1}^{L} N_h$: Tamanho da população;

- $W_h = \frac{N_h}{N}$: Peso do estrato, com $\sum_{h=1}^{L} W_h = 1$;
- $f_h = \frac{n_h}{N_h}$: Fração amostral do estrato;
- $\overline{Y}_h = \frac{\sum_{i=1}^{L} y_{hi}}{N_h}$: Média populacional do estrato;
- $\overline{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$: Média amostral do estrato;
- $S_h^2 = \sum_{i=1}^{n} \frac{(y_{hi}-\overline{y}_h)^2}{N_h-1}$: Variância populacional do estrato;
- $T_h = \sum_{i=1}^{N_h} y_{hi}$: Total do estrato;
- $\overline{Y} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi} = \frac{1}{N}\sum_{h=1}^{L} N_h\overline{Y}_h = \sum_{h=1}^{L} W_h\overline{Y}_h$: Média populacional;
- $\overline{y} = \frac{1}{n}\sum_{i=1}^{L} n_h\overline{y}_h$: Média amostral;
- $n = \sum_{h=1}^{L} n_h$: Tamanho da amostra;
- $s_h^2 = \sum_{i=1}^{n_h} \frac{(y_h-\overline{y}_h)^2}{n_h-1}$: Variância amostral do estrato;
- $\overline{y}_{es} = \overline{y}_{st} = \sum_{i=1}^{L} \frac{N_h\overline{y}_h}{N} = \sum_{h=1}^{L} W_h\overline{y}_h$, onde $N = N_1 + \ldots + N_h$

Suponha que o total da população de $N$ unidades seja particionado em $L$ estratos com $n_h$ unidades no $h$-ésimo estrato, onde $h = 1, 2, \ldots, L$. Seja a unidade, $h_i$, a $i$-ésima unidade no $h$-ésimo estrato associado com o valor de $y_{st}$. Nesse caso, obtem-se a média e a variância da $AE$ para uma amostra sorteada por um processo $AAS_C$ em que $\widehat{\mu}_h = \overline{y_h}$.

$$E[\overline{y}_{st}] = E\left[\sum_{h=1}^{L} W_h\overline{y}_h\right] = E\left(\sum_{h=1}^{L}\frac{N_h}{N}\overline{y}_h\right) = \frac{1}{N}\sum_{h=1}^{L} N_h E(\overline{y}_h) = \frac{1}{N}\sum_{h=1}^{L} N_h\overline{Y}_h = \overline{Y} \tag{1.42}$$

$$var[\overline{y}_{st}] = var\left[\sum_{h=1}^{L} W_h\overline{y}_h\right] = \sum_{h=1}^{L} W_h^2 var(\overline{y}_h) = \sum_{h=1}^{L} W_h^2(1-f_h)\frac{S_h^2}{n_h} \tag{1.43}$$

Um estimador para (1.43) é dado por

$$\widehat{var}[\overline{y}_{st}] = \sum_{h=1}^{L} W_h^2(1-f_h)\frac{s_h^2}{n_h} \tag{1.44}$$

e o total populacional é

$$T_{st} = \sum_{h=1}^{L} N_h \overline{y}_h \tag{1.45}$$

A variância para o total populacional (1.45) e o estimador dessa variância é dada por

$$var[\widehat{T}_{st}] = \sum_{h=1}^{L} N_h^2 Var[\overline{y}_h] = \sum_{h=1}^{L} N_h^2 \frac{S_h^2}{n_h} \tag{1.46}$$

$$\widehat{var}(\widehat{T}_{st}) = \sum_{h=1}^{L} N_h^2 (1-f) \frac{s_h^2}{n_h} \tag{1.47}$$

Nesse desenho amostral, há dois fatos predominantes para a determinação do número de estratos $L$ ao se realizar uma $AE$. O primeiro é saber a que taxa a variância de $\overline{y}_{st}$ decresce quando $L$ cresce. O segundo é saber quanto custo de um levantamento será afetado quando $L$ sofrer um aumento.

Essa distribuição das $n$ unidades da amostra pelos estratos chama-se alocação da amostra, que garante a precisão do procedimento amostral (Scheaffer et al., 1996). Isso significa que quanto maior a variância do estrato, maior deve ser o tamanho da amostra, $n_h$; mas isso deve ser balanceado com o tamanho do estrato, $W_h$. No caso da $AE$, a alocação pode ser de três tipos: proporcional, uniforme e ótima. Isso será feito supondo que dentro de cada estrato foi realizado uma $AAS_C$.

Com a alocação proporcional ($AE_{pr}$), como o próprio nome diz, a amostra de tamanho $n$ é distribuída proporcionalmente ao tamanho dos estratos e tem-se $n_h = W_h \times n = \frac{N_h}{N} \times n$. Dessa forma, $\overline{y}_{st} = \sum_{h=1}^{L} W_h \overline{y}_h = \overline{y}$, $f_h = \frac{n_h}{N_h} = \frac{n \times W_h}{N W_h} = \frac{n}{N}$ e $\frac{W_h^2}{n_h} = \frac{W_h^2}{n W_h} = \frac{W_h}{n}$.

Nesse caso, a variância da média e do total populacional são:

$$var(\overline{y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{S_h^2}{n_h} \tag{1.48}$$

$$var(\widehat{T}) = \frac{N-n}{N} \sum_{h=1}^{L} N_h S_h^2 \tag{1.49}$$

De (1.43), e sabendo que $s_h^2$ é um estimador não-viesado para $S_h^2$ (como provado

em (1.17) e (1.18)), tem-se que

$$V_{pr} = \widehat{var}(\overline{y}_{st}) = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{\frac{N_h}{N}n}{N_h}\right) \frac{s_h^2}{\frac{N_h}{N}n} = \sum_{h=1}^{L} \frac{N_h^2}{N^2} \left(\frac{NN_h - N_h n}{NN_h}\right) \frac{s_h^2}{N_h} \frac{N}{n}$$
$$= \sum_{h=1}^{L} \frac{N_h}{N} \left(\frac{N-n}{N}\right) \frac{s_h^2}{n} = \frac{(1-f)}{n} \sum_{h=1}^{L} W_h s_h^2 \tag{1.50}$$

No caso da alocação uniforme ($AE_{un}$), o tamanho de cada estrato é uniforme, ou seja, é o mesmo em cada estrato. Com isso, indica-se esse tipo de procedimento para quem deseja obter estimativas separadas para cada estrato. Assim, para cada estrato $L$, $n_h = \frac{n}{L} = b$ e $f_h = \frac{b}{N_h}$ (Bolfarine e Bussab, 2005). Dessa forma, é fácil verificar que $\overline{y}_{st}$ também é um estimador não-viesado para a média populacional. A sua variância e seu estimador de variância, bem como o total populacional são dados por (Cochran, 1977)

$$V_{un} = var(\overline{y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{S_h^2}{b} \tag{1.51}$$

$$var(\widehat{T}) = \sum_{h=1}^{L} N_h(N_h - n_h) \frac{S_h^2}{n_h} \tag{1.52}$$

$$\widehat{var}(\overline{y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{s_h^2}{b} \tag{1.53}$$

No caso da alocação ótima ($AE_{ot}$), o valor de $n$ pode ser selecionado de acordo com dois aspectos: primeiro, seleciona-se $n$ para minimizar a $var(\overline{y}_{st})$ para um custo específico de uma amostra; segundo, para minimizar o custo para um valor específico de $var(\overline{y}_{st})$.

Seja $C$, o custo total da amostragem; $c_0$, o custo fixo da amostragem; $c_h$, o custo da unidade amostral no estrato $h$, onde geralmente esses dois últimos valores são conhecidos. Essa função custo é expressa por:

$$C = c_0 + \sum_{h=1}^{L} c_h + n_h \tag{1.54}$$

Assim, em uma $AE_{ot}$ com uma função de custo linear, define-se o valor de $n_h$ de forma que a variância da média estimada $\overline{y}_{st}$ seja mínima para um custo fixo $c$ e o

custo seja mínimo para uma variância fixa $v$. Isso é verificado quando $n_h \propto \frac{W_h S_h}{\sqrt{C_h}}$.

Dessa forma, obtem-se as variações para o valor de $n_h$: O valor mínimo ocorre quando $n_h = \frac{nW_h \frac{S_h}{\sqrt{c_h}}}{\sum_{h=1}^{L} W_h \frac{S_h}{\sqrt{c_h}}}$, mas o valor de $n$ é determinado conforme a não variação de $c$ e $v$. Se o custo for fixo, $n = \frac{(c-c_0)\sum_{h=1}^{L} W_h \frac{S_h}{\sqrt{c_h}}}{\sum_{h=1}^{L} W_h S_h \sqrt{c_h}}$; mas se $v$ for fixo $n = \frac{\left(\sum_{h=1}^{L} W_h S_h \sqrt{c_h}\right)\left(\sum_{h=1}^{L} W_h \frac{S_h}{\sqrt{c_h}}\right)}{v+\frac{1}{n}\sum_{h=1}^{L} W_h S_h^2}$ (Cochran, 1977).

Ainda, se $c_h = c$, tem-se que $C = c_0 + c_n$, caracterizando uma alocação ótima para um tamanho de amostra $n$ fixo. Dessa maneira, obtem-se a alocação ótima de Neyman quando dado um $n$ fixo, a $var(\overline{y}_{st})$ é minimizada se $n_h = \frac{nW_h S_h}{\sum_{h=1}^{L} W_h S_h} = \frac{nN_h S_h}{\sum_{h=1}^{L} N_h S_h}$. Porém, se os custos e as variâncias são iguais, a alocação ótima é a proporcional, onde $n_h = \frac{nN_h S_h}{S_h \sum_{h=1}^{L} N_h} = \frac{nN_h}{N} = nW_h$ e $var(\overline{y}_{st}) = \frac{\left(\sum_{h=1}^{L} W_h S_h\right)^2}{n} - \frac{\sum_{h=1}^{L} W_h S_h^2}{n}$ (Bolfarine e Bussab, 2005).

Seja $\sigma^2$, a variância populacional, $\sigma_d^2 = \sum_{i=1}^{L} W_h \sigma_h^2$, a variância dentro do estrato e $\sigma_e^2 = \sum_{h=1}^{L} W_h(\mu_h - \mu)^2$, a variância entre os estratos. Observa-se que quando todos os estratos têm a mesma média, $\mu_h$, onde $h = 1, \ldots, L$, a variância populacional $\sigma^2 = \sigma_d^2 + \sigma_e^2$ coincide com $\sigma_d^2$. Outro ponto é que quanto maior for $\sigma_e^2$, maior é a diferença entre $\sigma^2 - \sigma_d^2$. Assim, $S^2 = \sum_{h=1}^{L} \frac{N_h - 1}{N-1} S_h^2 + \sum_{h=1}^{L} \frac{N_h}{N-1}(\mu_h - \mu)^2 \simeq \sigma_d^2 + \sigma_e^2$ (Bolfarine e Bussab, 2005). Determina-se, entã,o a variância de cada alocação,

$$V(AAS_C) = var(\overline{Y}) = \frac{\sigma^2}{n} = \frac{\sigma_d^2}{n} + \frac{\sigma_e^2}{n} = V_{pr} + \frac{\sigma_e^2}{n} \tag{1.55}$$

$$V(AE_{pr}) = \frac{1}{n}\sum_{h=1}^{L} W_h \sigma_h^2 = \frac{\sigma_d^2}{n} \Rightarrow V_{AAS_C} \geq V_{pr} \tag{1.56}$$

$$V(AE_{ot}) = \frac{1}{n}\left(\sum_{h=1}^{L} W_h \sigma_h\right)^2 \Rightarrow V_{ot} \leq V_{pr} \tag{1.57}$$

Assim, com relação á $AAS_C$, tem-se que $V_{ot} < V_{pr} < V_{AAS_C}$. Como consequência, quanto mais heterogêneo são os dados, mais indica-se fazer o uso da alocação ótima. Todavia, sempre que os estratos tiverem médias distintas, $\sigma_e^2$ grande, deve-se usar alocação proporcional ou ótima. Além disso, se os desvios padrões de cada estrato

forem discrepantes entre si, $\sigma_{dp}$ grande, é indicado usar alocação ótima. Logo,

$$V_{pr} - V_{ot} = \frac{1}{n}\left[\sum_{h=1}^{L} W_h\sigma_h^2 - \left(\sum_{h=1}^{L} W_h\sigma_h\right)^2\right] = \frac{1}{n}\sum_{h=1}^{L} W_h\left(\sigma_h - \overline{\sigma}\right)^2 = \frac{\sigma_{dp}^2}{n} \quad (1.58)$$

Semelhantemente à (1.41), pode-se obter o $EPA$ entre um planejamento de $AE$ com alocação proporcional, de um com alocação ótima e a de um $AAS_C$.

$$Deff(AE_{pr}) = \frac{V_{pr}}{V_{AAS_C}} = \frac{\sigma_{dp}^2}{\sigma^2} = 1 - \frac{\sigma_e^2}{\sigma^2} \quad (1.59)$$

$$Deff(AE_{ot}) = \frac{V_{ot}}{V_{AAS_C}} = \frac{\sigma_e^2 - \sigma_{dp}^2}{\sigma^2} \quad (1.60)$$

$$Deff(AE_{AAS_C}) = \frac{V_{AE_C}}{V_{AAS_C}} = \sum_{h=1}^{L} W_h\frac{W_h}{w_h}\left(\frac{\sigma^2}{\sigma}\right)^2 \quad (1.61)$$

Analogamente, pelo TCL sabe-se que para $n$ suficientemente grande, tem-se que

$$\frac{\overline{y}_{st} - \mu}{\sqrt{\sum_{h=1}^{L} W_h^2\frac{\sigma_h^2}{n_h}}} \sim N(0,1) \quad (1.62)$$

$$\frac{\widehat{\tau_{st}} - \tau}{\sqrt{\sum_{h=1}^{L} N_h^2\frac{\sigma_h^2}{n_h}}} \sim N(0,1) \quad (1.63)$$

e os intervalos de confiança aproximados para $\overline{y}_{st}\ \epsilon\ \tau$, ou seja,

$$P\left(\overline{y}_{st} - z_{\alpha/2}\sqrt{\sum_{h=1}^{L} W_h^2\frac{s_h^2}{n_h}} \leq \mu \leq \overline{y} + z_{\alpha/2}\sqrt{\sum_{h=1}^{L} W_h^2\frac{s_h^2}{n_h}}\right) \simeq 1 - \alpha \quad (1.64)$$

Para se determinar o tamanho da amostra $n$, é relevante saber o tipo da alocação e se será considerado ou não o *fpc*. Assim, o tamanho amostral geral é dada por $n = \frac{\sum_{h=1}^{L}\frac{W_h^2S_h^2}{w_h}}{\left(\frac{d}{t}\right)^2 + \frac{1}{N}\sum_{h=1}^{L} W_hS_h^2}$, onde a variância mínima é $V_{min}(\overline{y}_{st}) = \frac{\sum_{h=1}^{L}(W_hS_h)^2}{n} - \frac{\sum_{h=1}^{L} W_hS_h^2}{N}$. Assim, pode-se adaptar o tamanho amostral para o caso desejado.

Devido ao fato da amostragem espacial adaptável fazer uso de conglomerados, a próxima Seção abordará a amostragem por conglomerado clássica.

## 1.5 Amostragem por Conglomerados

A amostragem por conglomerados ($AC$) é realizada ao se dividir a população em subpopulações distintas, conglomerados (ou do inglês *clusters*); como bairros, prédios, famílias. Alguns desses conglomerados são selecionados segundo a $AAS$ e todos os indivíduos nos conglomerados selecionados são observados (Bolfarine e Bussab, 2005). Com isso, usa-se a $AC$ quando em algumas situações não existem um cadastro dos elementos da população, mas sim de seus conglomerados; ou mesmo quando há esses cadastros, mas é mais econômico esse tipo de amostragem.

Assim, diferentemente da $AE$, nesse plano amostral deseja-se maior heterogeneidade nos conglomerados. Porém, esse fato pode se tornar um inconveniente dado que na $AC$ as unidades dentro de um mesmo conglomerado tendem a ter valores parecidos, o que torna esse plano menos eficiente do que a $AAS$ e do que $AE$; mas, por outro lado, é mais econômico (Scheaffer et al., 1996). Em síntese, a $AC$ tende a ter: um menor custo; maior variância; maiores problemas para análises estatísticas, pois $n_i$ é uma variável aleatória (v.a.) e é preciso saber qual é a sua distribuição de frequência.

A $AC$ pode ser realizada em um, dois ou mais estágios, com probabilidades iguais ou desiguais com ou sem reposição e, ainda, com conglomerados de igual ou diferentes tamanhos. Inicia-se, então, o estudo da amostragem em um único estágio com probabilidades iguais para conglomerados de mesmo tamanho.

### 1.5.1 Amostra aleatória com probabilidades iguais

Nesse plano de $AC$, os conglomerados serão sorteados por meio de uma $AAS_C$ (conforme Seção 1.3.1) e dentro de cada um desses conglomerados serão entrevistados todos os indivíduos.

#### 1.5.1.1 Conglomerados de tamanhos iguais

Seja $P = \{C_1, C_2 \ldots, C_N\}$, a população; $C_i = \{E_{i1}, E_{i2}, \ldots, E_{iM}\} = \{Y_{i1}, Y_{i2}, \ldots, Y_{iM}\}$, os $Y_{iM}$ elementos de cada conglomerado; $A = \{c_1, c_2, \ldots, c_n\}$, a amostra onde $c_i = \{y_{i1}, y_{i2}, \ldots, y_{iM}\}$; $nM$, o tamanho da amostra. Define-se, então:

- $\overline{\overline{Y}} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} Y_{ij}$, média da população;
- $\overline{Y}_i = \frac{1}{n}\sum_{j=1}^{M} Y_{ij}$, média do conglomerado $i$;
- $S^2 = \frac{1}{NM-1}\sum_{i=1}^{N}\sum_{j=1}^{M}(Y_{ij} - \overline{\overline{y}})^2$, variância populacional;
- $S_i^2 = \frac{1}{n-1}\sum_{j=1}^{M}(Y_{ij} - \overline{Y}_i)^2$, variância do conglomerado $i$;
- $T_i = Y_i = \sum_{j=1}^{M} Y_{ij}$, total do conglomerado $i$;
- $S_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{Y}_i - \overline{\overline{y}})^2$, variância da média do conglomerado $i$;
- $\overline{\overline{y}} = \frac{1}{nM}\sum_{i=1}^{n}\left(\sum_{j=1}^{m} y_{ij}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j=1}^{m} Y_{ij}\right) = \frac{1}{n}\sum_{i=1}^{n}\overline{y}_i$, estimador não-viesado para $\overline{\overline{y}}$.

Sabe-se que de uma $AAS$ de conglomerados, cada um contendo $n$ elementos selecionados de uma população de $N$ conglomerados, então a média amostral por elemento, $\overline{\overline{y}}$, é uma estimativa não-viesada de $\overline{\overline{Y}}$ com variância dada por:

$$V(\overline{\overline{y}}) = \frac{(1-f)}{n}\frac{N(M-1)}{M^2(N-1)}S^2[1+(M-1)\rho] = \frac{(1-f)}{nM}S^2[1+(M-1)\rho] \quad (1.65)$$

onde $\rho$ é o coeficiente de correlação intraclasse que varia entre $\frac{-1}{M-1} \leq \rho \leq 1$. Esse coeficiente é menos preciso para um determinado tamanho de amostra se $\rho > 0$; mas se existe heterogeneidade interna no conglomerado, $\rho < 0$, e a $AC$ é mais eficiente do que a $AAS$. O $\rho$ é dado por (Cochran, 1977):

$$\rho = \frac{E(Y_{ij} - \overline{\overline{Y}})(Y_{ik} - \overline{\overline{Y}})}{E(Y_{ij} - \overline{\overline{Y}})^2} = \frac{2\sum_{i=1}^{N}\sum_{i\neq j<k}\sum(Y_{ij} - \overline{\overline{Y}})(Y_{ik} - \overline{\overline{Y}})}{(M-1)(NM-1)S^2} \quad (1.66)$$

Nesse caso, pode-se fazer o uso da estimativa de uma proporção de um conglomerado com $M$ elementos, onde $p_i = \frac{a_i}{M}$ seja a proporção de elementos do conglomerado de ordem $i$ que pertence à categoria $C$. Assim, seleciona-se uma $AAS$ de $n$ conglomerados, média $p$ das $p_i$ observações amostradas. Dessa forma, obtem-se

$$V(p) = \frac{(1-f)}{n}\frac{\sum_{i=1}^{N}(p_i - p)^2}{N-1} \quad (1.67)$$

$$\widehat{V}(p) = \frac{(1-f)}{n}\frac{\sum_{i=1}^{N}(p_i - \overline{p})^2}{M-1} \tag{1.68}$$

onde $\overline{p} = \frac{\sum_{i=1}^{n} p_i}{n}$.

Mas se for tomado uma $AAS_C$ de $nM$ elementos, a variância de $p$ é dada por:

$$V(p)_{Bin} = \left(\frac{N-M}{N-1}\right)\frac{PQ}{n} = \left(\frac{MN-nM}{N-1}\right)\frac{PQ}{nM} \approx \left(\frac{N-M}{N}\right)\frac{PQ}{nM} \tag{1.69}$$

onde essa última aproximação é válida para um valor de $N$ grande.

Consequentemente, calculando o *Deff* entre essas variâncias tem-se:

$$\frac{V(p)}{V(p)_{Bin}} = \frac{\left(\frac{N-n}{N^2 n}\right)\sum_{i=1}^{n}(p_i - p)^2}{\frac{N-n}{N}\frac{PQ}{nM}} \approx \frac{\sum_{i=1}^{n}(p_i - p)^2}{NPQ} \tag{1.70}$$

onde essa última aproximação é válida para um valor de $N$ grande.

#### 1.5.1.2 Conglomerados de tamanhos desiguais

Seja $P = \{C_1, C_2 \ldots, C_N\}$, a população; $M_i$, os números de elementos do conglomerado $i$; $C_i = \{E_1, E_2, \ldots, E_{M_i}\} = \{Y_1, Y_2, \ldots, Y_{M_i}\}$, onde $Y_i = T_i = \sum_{j=1}^{M_i} Y_{ij}$, o total do conglomerado $i$; $A = \{c_1, c_2, \ldots, c_n\}$, a amostra onde $c_i = \{y_1, y_2, \ldots, y_{M_i}\}$; $nM$, o tamanho da amostra; $\overline{Y}_i = \frac{\sum_{j=1}^{M_i} Y_{ij}}{M_i}$, a média do conglomerado $i$. Pode, então, ser calculado a $AAS$ de $n$ clusters, bem como a $AAS$ de $n$ elementos com estimadores tipo razão.

No primeiro caso, obtem-se que a média de $\overline{y} = \frac{\sum_{i=1}^{n}}{n} y_i$ tem como estimador não-viesado o $\widehat{Y} = \widehat{T} = N \times \overline{y}$, pois $E(\widehat{Y}) = Y$ e $E(\overline{y}) = \overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$. Assim, obtem-se

$$var(\overline{y}) = \frac{N-n}{N}\frac{S_y^2}{n} = \frac{(1-f)\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N-1} \tag{1.71}$$

$$\widehat{var}(\overline{y}) = \frac{(1 - f\sum_{i=1}^{n}(y_i - \overline{y})^2)}{n(n-1)} \tag{1.72}$$

$$\widehat{var}(\widehat{Y}) = N^2\widehat{var}(\overline{y}) \tag{1.73}$$

Para o segundo caso tem-se $M_0 = \sum_{i=1}^{N} M_i$ o número total de elementos na população como valores conhecidos. Com isso, $M_i$ é tomado como uma variável auxiliar de $X_i$. Dessa forma, tem-se $\widehat{T}_{YR} = \widehat{Y}_R = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{N} M_i} = \widehat{R} \times M_0$; mas $\widehat{R} = \frac{Y}{X} =$

$\frac{Y}{M_0} = \overline{\overline{Y}}$, onde $\overline{\overline{Y}} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M_i} Y_{ij}}{M_0} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M_i} Y_{ij}}{\sum_{i=1}^{N} M_i} = \frac{Y}{M_0}$. Além disso,

$$var(\widehat{Y}_R) = \frac{N^2(1-f)\sum_{i=1}^{N}(Y_i - M_i\overline{\overline{Y}})^2}{n(N-1)} = \frac{N^2(1-f)\sum_{i=1}^{N} M_i^2(Y_i - \overline{\overline{Y}})^2}{n(N-1)} \quad (1.74)$$

$$\widehat{var}(\widehat{Y}_R) = \frac{N^2(1-f)\sum_{i=1}^{n}(y_i - M_i\overline{\overline{y}})^2}{n(n-1)} \quad (1.75)$$

onde $\overline{\overline{y}} = \frac{\widehat{Y}}{M_0} = \frac{\sum_{i=1}^{n} y_i N}{nM_0}$ e $\overline{\overline{y}}_R = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i}$, a média amostral por elemento.

Porém, em alguns tipos de amostragem, como será o caso da amostragem espacial adapável do próximo Capítulo, cada unidade pode ter probabilidades diferentes ao ser adicionada na amostra (Thompson e Seber, 1996). Seja $\Pi_i$, essa probabilidade de inclusão; $E[\overline{y}] = \mu$; $Z = \sum_{i=1}^{N} z_i I_i$ e $E[Z] = \sum_{i=1}^{N} z_i EI_i = \sum_{i=1}^{N} z_i \Pi_i$ então,

$$E[\overline{y}] = \frac{1}{n}\sum_{i=1}^{N} y_i P(I_i = 1) = \frac{1}{n}\sum_{i=1}^{N} y_i \Pi_i. \quad (1.76)$$

Esse cálculo mostra que $E[\overline{y}] \neq \mu$. Assim, $\overline{y}$ é um estimador viesado. Isso pode ser facilmente verificado para $\widehat{Y}_R$, ou seja, $E(\widehat{Y}_R) \neq Y$. Para solucionar o viés desses estimadores nesses casos de probabilidades de inclusão tem-se dois estimadores para o caso de amostragem com probabilidade desigual: Hansen-Hurwitz, no caso com reposição e Horvitz-Thompson, no caso sem reposição.

Nesse trabalho será abordado o estimador de Horvitz-Thompson por ser uma generalização do outro estimador e como foi demonstrado na Seção 1.3, a diferença entre o caso com e sem reposição é basicamente dada pela devolução ou não do elemento na amostra. Além disso, esse estimador modificado será usado no cálculo dos estimadores para o caso de dados raros.

### 1.5.2 Amostra aleatória com probabilidades desiguais e sem reposição

Nesta Seção, assume-se que para calcular o estimador Horvitz-Thompson as unidades participantes da amostra são selecionadas sem reposição. Seja $\Pi_i = P(I_i = 1) = E[I_i]$ a probabilidade do modelo onde a unidade $i$ está inserida na amostra e

que $\Pi_i$ é estritamente maior do que zero para cada unidade na população. Assim, o estimador Horvitz-Thompson de $\mu$ é dado por (Thompson e Horvitz, 1952)

$$\widehat{\mu}_{HT} = \frac{1}{N}\sum_{i=1}^{n}\left(\frac{y_i}{\Pi_i}\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i I_i}{\Pi_i} \tag{1.77}$$

onde $y_1, y_2, \ldots, y_n$ representam os valores de $y$ das $n$ legendas distintas na amostra, $I_i$ é a função indicadora ligada à inclusão da unidade $i$ na amostra e que a amostragem pode ser com ou sem reposição. Caso os valores de $\Pi_i$ sejam diferentes, então a legenda do estimador será dependente, dado que depende no conhecimento específico de qual legenda será amostrada.

Seja $\widehat{\mu}_{HT} = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i I_i}{\Pi_i}$ então, obtem-se um estimador não-viesado para a média:

$$E[\widehat{\mu}_{HT}] = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{\Pi_i}E[I_i] = \frac{1}{N}\sum_{i=1}^{N} y_i = \mu \tag{1.78}$$

Assim, $\widehat{\mu}_{HT}$ é não-viesado.

De (1.1), (1.2) e (1.4) tem-se que

$$Var[\widehat{\mu}_{HT}] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i=1}^{N}\left(\frac{\Pi_{ij} - \Pi_i\Pi_j}{\Pi_i\Pi_j}\right)\frac{y_i}{\Pi_i}E[I_i] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i=1}^{N}\left(\frac{\Pi_{ij} - \Pi_i\Pi_j}{\Pi_i\Pi_j}\right) y_i y_j \tag{1.79}$$

Assim, para o caso de $AAS_S$, onde $\Pi_i = \frac{n}{N}$ e $\Pi_{ij} = \frac{n(n-1)}{N(N-1)}$,

$$var[\overline{y}] = \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right) \tag{1.80}$$

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

$$\widehat{var\overline{y}} = \frac{s^2}{n}\left(1 - \frac{n}{N}\right) \tag{1.81}$$

onde $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$; e para o caso de $AAS_C$, onde a probabilidade de inclusão é $\Pi_i = 1 - (1-p_i)^n$ e $\Pi_{ij} = (1-p_i)^n + (1-p_j)^n - [1 - (1-p_i-p_j)^n]$, basta substituir os valores de $\Pi_i$ e de $\Pi_{ij}$ nas Equação (1.79).

A técnica a ser apresentada a seguir é chamada de subamostragem ou amostra-

gem em dois estágios, devido a *Mahalanobis*.

### 1.5.3 Amostragem por Conglomerados em Dois Estágios

Após a população estar agrupada em $n$ conglomerados, descreve-se o plano amostral do seguinte modo (Cochran, 1977):

- Seleciona-se uma amostra de unidades, os conglomerados, denominadas de unidades primárias, UPS (ou do inglês *PSU - Primary Sampling Unit*);
- Depois disso, seleciona-se uma amostra de subunidades, chamadas unidades secundárias de seleção, USS (ou do inglês *SSU - Secondary Sampling Unit*) de cada uma das UPS selecionadas.

Dessa forma, ao supor que cada conglomerado na população seja dividido em um número menor de unidades, ou seja, seleciona-se uma amostra de $n$ conglomerados; obtem-se um aumento de eficiência, pois o tamanho da amostra permanece o mesmo. Assim, essas subamostras podem ser combinadas com qualquer tipo de amostragem da UPS como a estratificação (Bolfarine e Bussab, 2005).

Vamos considerar inicialmente que todas as unidades tem o mesmo número de elementos, $M$ e de cada conglomerado na amostra do primeiro estágio, seleciona-se uma subamostra de $m$ elementos. A vantagem dessa amostragem é que essa é mais flexível do que em estágio único e se igualam quando $M = m$.

Assim, o tamanho da amostra final é dado por $f = \frac{nm}{NM}$; e no segundo estágio o número de amostras possíveis são $(C_m^M)^n$, enquanto que no total são possíveis $C_n^N(C_m^M)^n$.

Um modo de se obter a média e a variância para esse tipo de amostragem é calcular medidas como a média da estimativa sobre todas as seleções do segundo estágio e então a média sobre todas as possíveis seleções de $n$ unidades pelo plano amostral. Isso pode ser feito por meio da esperança de uma estimativa $\theta$. Logo,

$$E[\widehat{\theta}] = E_1[E_2(\widehat{\theta})] \tag{1.82}$$

$$var(\widehat{\theta}) = V_1[E_2(\widehat{\theta})] + E_1[V_2(\widehat{\theta})] \tag{1.83}$$

ou seja, usa-se a propriedade de esperança matemática onde $E[E(X|Y)] = E(X)$. Dessa forma, a esperança da estimativa $\theta$ no segundo estágio dado que se tem uma amostra no primeiro estágio é igual a esperança da estimativa da mostra do segundo estágio. Dessa forma, para processos com $n = 1$, amostragem em dois estágios, tem-se

- $\overline{\overline{Y}} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} Y_{ij} = \frac{1}{N}\sum_{i=1}^{N}\overline{Y}_i$, a média global da população;
- $y_{ij}$, o valor obtido para o elemento de ordem $j$, na unidade primária $i$;
- $\overline{y}_i = \sum_{j=1}^{m}\frac{y_{ij}}{m}$, o valor médio amostral por elemento, na unidade primária de ordem $i$;
- $\overline{\overline{y}} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} = \sum_{i=1}^{n}\frac{\overline{y}_i}{n}$, o valor médio global da amostra, por elemento;
- $S_1^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{Y}_i - \overline{\overline{Y}})^2$, a variância entre as UPS, ou seja, entre os valores médios das unidades primárias;
- $S_2^2 = \frac{1}{N(M-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}(Y_{ij} - \overline{Y}_i)^2$, a variância entre as USS dentro das UPS, ou seja, é a variância entre os elementos dentro das unidades primárias.

Com isso, a média estimada e a variância da média estimada é obtida de forma que as $n$ unidades primárias e as $m$ unidades secundárias de cada UPS escolhidas são selecionadas por $AAS$, isto é,

$$\begin{aligned} E\left(\overline{\overline{y}}\right) &= E_1\left[E_2\left(\overline{\overline{y}}\right)\right] = E_1\left[E_2\left(\frac{1}{n}\sum_{i=1}^{n}\overline{y}_i\right)\right] = E_1\left[\frac{1}{n}\sum_{i=1}^{n}E_2(\overline{y}_i)\right] = E_1\left(\frac{1}{n}\sum_{i=1}^{n}\overline{Y}_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}E_1(\overline{Y}_i) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{N}\sum_{j=1}^{M}\overline{Y}_j\right) = \frac{1}{n}n\frac{1}{N}\sum_{j=1}^{N}\overline{Y}_j = \overline{\overline{Y}} \end{aligned} \tag{1.84}$$

$$Var(\overline{\overline{y}}) = \left(\frac{N-n}{N}\right)\frac{S_1^2}{n} + \left(\frac{M-m}{M}\right)\frac{S_2^2}{mn} = (1-f_1)\frac{S_1^2}{n} + (1-f_2)\frac{S_2^2}{mn} \tag{1.85}$$

onde $f_1 = \frac{n}{m}$ e $f_2 = \frac{m}{M}$ e o estimador da variância (1.85) é dada por

$$\widehat{var}(\overline{\overline{y}}) = (1-f_1)\frac{s_1^2}{n} + f_1(1-f_2)\frac{s_2^2}{mn} \tag{1.86}$$

onde $s_1^2 = \sum_{i=1}^{n} \frac{(\overline{y}_i - \overline{\overline{y}})^2}{n-1}$ e $s_2^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(y_{ij} - \overline{y}_i)^2}{n(m-1)}$.

Considera-se, então, uma amostragem estratificada para as UPS em uma amostra de dois estágios e supõem-se que essas unidades são constantes, ou seja, o tamanho das UPS são constantes, para um determinado estrato, mas podem variar de estrato para estrato.

### 1.5.4 Amostragem Estratificada dos Conglomerados

Seja o estrato $h$ que contém $N_h$ UPS, cada uma com $M_h$ USS e os tamanhos amostras são dados por, $n_h$ e $m_h$, respectivamente. Então, sabe-se que o valor médio estimado da população por unidade de segundo estágio é dado por (Cochran, 1977):

$$\overline{\overline{y}}_{st} = \frac{\sum_{h=1}^{L} N_h M_h \overline{\overline{y}}_h}{\sum_{h=1}^{L} N_h M_h} = \sum_{h=1}^{L} W_h \overline{\overline{y}}_h \tag{1.87}$$

onde tem-se que $W_h = \frac{N_h M_h}{\sum_{h=1}^{L} N_h M_h}$, refere-se ao estrato de unidades de segundo estágio; e $\overline{\overline{y}}_h = \frac{\sum_{i=1}^{n} \overline{y}_{hi}}{n_h}$ é o valor médio da amostra nesse estrato.

Além disso, tem-se a variância e a estimativa amostral da média (1.87) dado, respectivamente, por (Cochran, 1977):

$$V(\overline{\overline{y}}_{st}) = \sum_{h=1}^{L} W_h^2 \left( \frac{1 - f_{1h}}{n_h} S_{1h}^2 + \frac{1 - f_{2h}}{n_h m_n} S_{2h}^2 \right) \tag{1.88}$$

onde $f_{1h} = \frac{n_h}{N_h}$ e $f_{2h} = \frac{m_h}{M_h}$

$$\widehat{V}(\overline{\overline{y}}_{st}) = \sum_{h=1}^{L} W_h^2 \left[ \frac{1 - f_{1h}}{n_h} s_{1h}^2 + \frac{f_{1h}(1 - f_{2h})}{n_h m_h} s_{2h}^2 \right] \tag{1.89}$$

Para o total da população, o valor de $V(\overline{\overline{y}}_{st})$ da Equação (1.88 ) ou $\widehat{V}(\overline{\overline{y}}_{st})$ da Equação (1.89) é dado ao se multiplicar essas Equações por $\left( \sum_{h=1}^{L} N_h M_h \right)^2$ (Cochran, 1977).

# Capítulo 2

# Amostragem Espacial Adaptável

## 2.1 Introdução

A maioria dos métodos discutidos na teoria de amostragem está limitada a modelos de amostragem onde a seleção das amostras pode ser feita antes do estudo. Por outro lado, um novo método de amostragem que faz o uso dos dados obtidos é a amostragem espacial adaptável. A base conceitual desta amostragem foi desenvolvida no final dos anos 60 (Basu, 2002) e sua teoria tem sido melhorada consideravelmente nos últimos anos (Thompson, 1990, 1991; Ramsey e Seber, 1992; Brown, 1994; Thompson e Seber, 1996; Brown, 1996; Salehi e Seber, 1997; Khan e Muttlak, 2002; Stein e Ettema, 2003; Sengupta e Sengupta, 2011; Jain e Chang, 2004; Thompson, 2011; yu).

Nesse novo desenho, em que o processo de seleção pode depender sequencialmente dos valores observados da variável de interesse, que são definidas como *unidades*, há ganhos de eficiência. Isso pode ser verificado ao se fazer um levantamento de uma planta rara, onde um botânico pode se sentir inclinado a experimentar mais intensamente uma área em que uma planta está localizada para ver se outras ocorrem em uma moita (Thompson, 2002); ou em uma região contendo espécies ameaçadas ou uma contaminação da terra (Thompson e Seber, 1996).

Esse método consiste em dividir a área da população a ser estuda em unidades espaciais por meio de uma grade regular, que geralmente são de mesmo tamanho. A partir disso, o número de unidades de interesse são contados por meio de uma seleção

como a $AAS$. Em seguida, os vizinhos dessa unidade são adicionados à amostra da seguinte maneira: se esses vizinhos também tiverem unidades de interesse, seleciona-se os novos vizinhos desse vizinho e assim por diante até atingir o total amostral; se não tiverem unidades, para a seleção nessa unidade selecionada. Mesmo que a unidade selecionada por meio da amostra inicial não tiver unidades, ela será mantida na amostra e fará parte da amostra total, como na Figura 4.1.

Para utilizar essa técnica, entretanto, novos estimadores devem ser implementados para garantir o não-enviesamento e a imparcialidade das estatísticas calculadas (Thompson e Seber, 1996). A abordagem convencional feita no Capítulo 1 fornece a base para a amostragem espacial adaptável que será abordada nesse Capítulo.

## 2.2 Vantagens e Definições da Amostragem Espacial Adaptável

Segundo Thompson (2002), existem duas vantagens principais para a amostragem espacial adaptável. A primeira vantagem é a capacidade de incorporar características da população para obter estimativas mais precisas de densidade populacional. Por exemplo, as populações de plantas e animais, minerais e combustíveis fósseis tendem a exibir tendências de agregação natural (Sengupta e Sengupta, 2011; yu).

Uma vantagem secundária dessa amostragem é um aumento no rendimento de observações importantes como o número de espécies em vias de extinção observado. Isso pode resultar em estimativas de parâmetros de melhor qualidade.

Outro aspecto importante é que para um dado tamanho de amostra e um custo, uma informação mais valiosa pode ser obtida por meio da amostragem espacial adaptável do que é possível com sistemas convencionais (Thompson e Seber, 1996). Isso se deve ao fato de que, uma vez que a localização e forma desses agregados naturais não podem muitas vezes ser previsto, a amostragem espacial adaptável pode fornecer uma maneira de aumentar dramaticamente a eficácia do projeto de amostragem quando os vizinhos da unidade são adicionados.

Assim, o objetivo principal dos desenhos de amostragem espacial adaptável é aproveitar o padrão espacial da população para obter medidas mais precisas da abundância da população. Em muitas situações, a amostragem espacial adaptável é muito mais eficiente para uma determinada quantidade de esforço do que as técnicas convencionais (Thompson, 2002).

Como em uma situação de amostragem usual de uma população finita (Hansen e Hurwitz, 1943), a população consiste em $N$ unidades com rótulos de $\{1, 2, \ldots, N\}$ e com variáveis de interesse $y = \{y_1, y_2, \ldots, y_N\}$. A amostra $n$ é um conjunto ou consequências de rótulos identificando as unidades selecionadas dentre as observadas. Os dados consistem em valores observados, $y$, associados com seus rótulos e o objeto de interesse é estimar a média populacional, $\mu$, ou o total dos valores de $y$, $\tau$. O modelo amostral é a função $p(n|y)$ atribuindo uma probabilidade a uma possível amostra $n$, onde essas probabilidades de seleção dependem dos valores $y$ da população.

Define-se a seguir alguns termos a serem utilizados (Thompson, 1990):

- Vizinhança ou *neighborhood*: É um conjunto de quadrados que serão adicionados na mesma amostra se esse quadrado da grade satisfizer a mesma condição. No caso a ser estudado, define-se vizinhança como sendo os quatro quadrados: acima, abaixo e aos lados (direita e esquerda) do quadrado inicial. A condição para adicionar as unidades dos vizinhos de interesse é dada por um intervalo ou conjunto $C$ na amplitude da variável de interesse. A unidade $i$ satisfaz essa condição se $y_i \epsilon C$, ou seja, a unidade satisfaz a condição se a variável de interesse, $y_i$ é maior ou igual a alguma constante c, isto é, $C = \{y : y \geq c\}$.

- Conglomerado ou *cluster*: Considerando a coleção de todas as unidades que são observadas sob o modelo como o resultado de uma seleção de $i$ unidades. Essa coleção pode consistir na união de várias vizinhanças. Um conglomerado consiste, ainda, somente de um ponto no quadrado da grade se esse quadrado não satisfaz a condição de vizinhança, isto é, se for uma unidade de borda (quadrado selecionado na amostragem espacial adaptável, mas sem pontos amostrais).

- Rede ou *Network*: Uma rede é um conjunto de conglomerados que consiste de todas as unidades do conglomerado que satisfaz a condição de vizinhança.

- Unidade de borda ou *Edge unit* : qualquer unidade que não satisfaça a condição mas que estão na vizinhança de alguma unidade que a satisfaz, ou seja, são todos os quadrados da grade que foram selecionados mas não possuem pontos amostrais.

Thompson (2002) define um *bairro*, como todas as unidades vizinhas (ou regiões) dentro de uma determinada dimensão, como a Figura 2.1 que representa o caso unidimensional. A unidade inicial é ilustrada em azul, as quatro unidades em preto para a direita, à cima, à baixo e à esquerda da unidade de azul estão em "proximidade" da unidade de azul. Essas quatro unidades de preto pertencem ao bairro da unidade azul ou não, caso não atendam ao critério estabelecido.

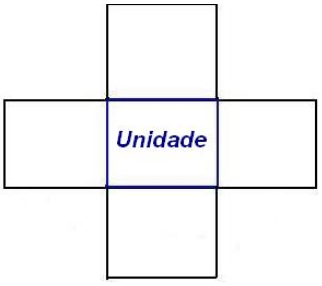


Figura 2.1: Unidade Inicial na Seleção da Amostragem Espacial Adaptável

Um outro conceito importante é que o bairro pode seguir qualquer padrão arbitrário, mas em relação a vizinhança prevalece a simetria: se a unidade $j$ está na vizinhança da unidade $i$, então a unidade $i$ está na vizinhança da unidade $j$ (Thompson, 2002). Isso significa que se a unidade fica em um bairro de unidades, então outras unidades são vizinhas dessa unidade.

Assume-se, então, que para cada unidade $i$ da população, uma vizinhança $A_i$ é definida, consistindo em um conjunto de unidades que incluem a unidade $i$. Essa vizinhança não depende dos valores $y$ da população (Thompson e Seber, 1996). No entanto, as unidades na vizinhança podem ser não contíguos (Thompson e Seber, 1996). Se qualquer uma dessas observações adicionais satisfazem a condição; então, em seguida, as observações de sua vizinhança são tomadas, ou seja, os vizinhos

das áreas selecionadas: superiores, inferiores, direita e esquerda são adicionados à amostra. Esse procedimento continua até que não haja observações que satisfazem a condição (Thompson, 1990).

Os tipos da amostragem espacial adaptável ocorrem quando (Thompson e Seber, 1996):

- Seleciona-se uma área da população e o processo descrito acima é iniciado, processo esse denominado de amostragem espacial adaptável por conglomerado (Thompson, 1990; Salehi e Seber, 1997; Khan e Muttlak, 2002);

- A população é dividida em estratos e o número de unidades são amostradas para cada estrato, chamado de amostragem espacial adaptável estratificada por conglomerados (Thompson, 1991);

- Uma espécie de elemento a ser analisado não estiver claramente visível na área determinada, então existe a possibilidade de esse elemento não ser detectado, mesmo que presente, denominado de detectabilidade incompleta (Thompson e Seber, 1994);

- Deseja-se analisar algumas regiões de uma população para verificar determinada característica, sem verificar todas as regiões, mas especialmente as que têm maiores níveis do elemento a ser estudado. Depara-se, então, com a amostragem espacial adaptável baseada em estatísticas de ordem (Thompson, 1996);

- Uma área é dividida em estratos e dentro desse o elemento é alocado aleatoriamente. Se a condição de parada for satisfeita não seleciona mais elementos; se não, usa-se a técnica de amostragem espacial adaptável. Isso significa que o conceito de vizinhos não é utilizado nesse caso, ou seja, em vez de usar a amostragem espacial adaptável localmente, aloca as amostras nos estratos adaptavelmente (Wald, 1947; Robbins, 1952; Zacks, 1970; Siegmund, 1985; Francis, 1991; Cochran, 1977; Thompson, 2002);

- Se a investigação envolver mais de uma espécie rara ao mesmo tempo, então fica caracterizado uma situação multivariada onde se tem um vetor de contagem em vez de uma única contagem para cada unidade da grade. Essa

amostragem pode ser aplicada tanto para número de animais como para medir as características desses animais. (Bethel; Kokan e Khan, 1967; Cochran, 1977; Thompson, 1993).

## 2.3 Amostragem Espacial Adaptável por Conglomerado

Observa-se que a amostragem espacial adaptável é realizada em amostragem por conglomerado, pois os dados a serem analisados são divididos em subpopulações distintas, como explicado na Seção 1.5 do Capítulo 1, conforme as coordenadas geográficas de uma determinada área. Dessa forma, o processo de amostragem espacial adaptável envolve a seleção de uma determinada área tendo registradas suas coordenadas geográficas e a verificação de dados nessa região, bem como a localização das unidades de interesse. Normalmente esses dados se encontram agrupados em uma determinada área conforme a Figura 2.2 (a), caracterizando a formação dos conglomerados.

Assim, tendo as coordenadas geográficas da área a ser analisada, primeiramente, desenha-se a grade nessa área para obter a localização espacial da área e dos dados de interesse a serem analisados, conforme indicado na Figura 2.2 (b). Depois, sorteia-se uma $AAS$ de grades e, em seguida, identifica-se as unidades de interesse (normalmente quando se encontra elementos dentro dessa grade) como na Figura 2.2 (c). Após isso, identifica-se sucessivamente os vizinhos dessas unidades selecionadas: superiores, inferiores, da direita e da esquerda, até esgotar o critério de seleção da amostra adaptável. Essa amostra adaptável é a amostra populacional ($n$) da grade regular, como na Figura 2.2 (d).

Em síntese, a amostragem espacial adaptável por conglomerado refere-se a modelos em que são feitos uma seleção por meio de um procedimento probabilístico como a $AAS$ e sempre que se verifica a variável de interesse em uma unidade selecionada, unidades adicionais na vizinhança dessa unidade são adicionadas à amostra (Thompson, 1990). Nos modelos a serem considerados, a amostra inicial pode ser

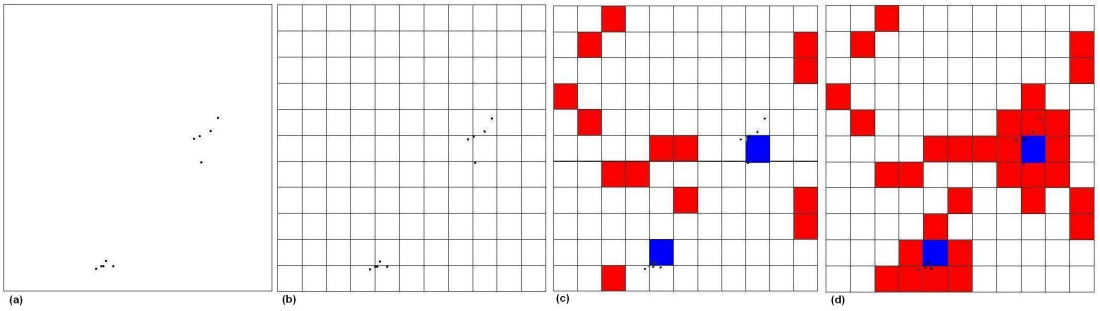


Figura 2.2: Passos da Amostragem Espacial por Conglomerados Adaptável

selecionada por meio de $AAS_C$ ou $AAS_S$.

### 2.3.1 Estimadores

Estimadores clássicos, como a média, são estimadores não-viesados da média da população sob um desenho amostral não adaptável, como ocorre na $AAS$, conforme visto nas Equações (1.13) e (1.32). Nessa seção serão abordados dois estimadores não-viesados para a amostragem espacial adaptável.

#### 2.3.1.1 Estimadores Usando a Probabilidade de Intersecção Inicial

Nessa seção será abordado um estimador baseado no estimador modificado de Horvitz-Thompson (Thompson, 1990) e será feita a sua comparação com a média da amostra inicial, dada pela primeira igualdade da Equação (1.12).

Quando uma amostra inicial de $n_1$ unidades é selecionada por uma $AAS_S$, essas unidades na amostra inicial são distintas devido a não reposição. Todavia, os próprios dados podem conter observações repetidas caso seja selecionado na amostra inicial mais de uma unidade no conglomerado. A unidade $i$ será incluída na amostra quando qualquer unidade da rede ao qual ela pertence, $A_i$, (incluindo ela) é selecionada como parte inicial da amostra; ou quando qualquer unidade da rede, em que a unidade $i$ é uma unidade de borda, é selecionada.

Seja $m_i$, o número de unidades na rede a que $i$ pertence; $a_i$, o número total de unidades da rede em que a unidade $i$ é uma unidade de borda. Note que se a unidade satisfizer o critério $C$, então $a_i = 0$; mas, se a unidade $i$ não satisfizer a condição,

então $m_i = 1$. A probabilidade de seleção da unidade $i$ em qualquer uma das $n_1$ observações é $p_i = \frac{m_i + a_i}{N}$. A probabilidade de a unidade $i$ estar inclusa na amostra é dada por (Thompson, 1990):

$$\Pi_i = P(I_i = 1) = 1 - \left[ \binom{N - m_i - a_i}{n_1} / \binom{N}{n_1} \right] \tag{2.1}$$

Quando a selação da amostra inicial for feita por $AAS_C$ observações repetidas nos dados podem ocorrer tanto por causa das possíveis seleções repetidas na amostra inicial quanto a seleção inicial de mais de uma unidade no conglomerado. Nesse desenho amostral, probabilidade de seleção da unidade $i$ em qualquer uma das $n_1$ observações é $p_i = \frac{m_i + a_i}{N}$ e a probabilidade de inclusão é dada pela Equação (**??**), $\alpha_i = 1 - (1 - p_i)^{n_1} = 1 - \left(1 - \frac{m_i + a_i}{N}\right)^n$ (Thompson, 1990).

Se os valores de $\Pi_i$ forem conhecidos para todas as unidades amostrais, poder-se-ia usar o estimador de Horvitz-Thompson dada pela Equação (1.77). Todavia, apesar de saber os valores de $m_i$ na Equação (2.1) para todas as unidades na amostra, apenas alguns valores de $a_i$ são conhecidos. Isso significa que seja $i$ uma unidade de borda em algum conglomerado pertencente à amostra, então todos os conglomerados que essa unidade está relacionada, não serão necessariamente amostrados. Com isso, o valor de $a_i$ será desconhecido. Para solucionar esse problema, adotou-se retirar o valor de $a_i$ na Equação (2.1) e considerar apenas uma inclusão de probabilidade parcial. Assim,

$$\Pi_i' = 1 - \left[ \binom{N - m_i}{n_1} / \binom{N}{n_1} \right] \tag{2.2}$$

Essa probabilidade, $\Pi_i'$ é agora considerada para $n$ redes em vez de conglomerados e pode ser entendida como a probabilidade de a unidade $i$ ser utilizada no estimador, isto é, a probabilidade de a amostra inicial interceptar $A_i$, a rede para a unidade $i$. Assim, obtem-se um estimador não-viesado baseado na probabilidade de intersecção inicial.

$$\widehat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i I_i'}{\Pi_i'} \tag{2.3}$$

onde $I_i'$ assume o valor 1 com probabilidade $\Pi_i'$ se a amostra inicial interceptar $A_i$, caso contrário assume o valor 0; e da Equação (1.77), $\widehat{\mu}_{HT} = \frac{1}{N} \sum_{i=1}^{n} \left( \frac{y_i}{\Pi_i} \right) =$

$\frac{1}{N}\sum_{i=1}^{N}\frac{y_i I_i}{\Pi_i}$, onde $y_1 \ldots y_n$ representam os $n$ valores distintos de unidades na amostra final e $I_i$ tem o valor 1 quando a unidade está incluída na amostra e 0 caso contrário.

Da Equação (1.3) e utilizando as propriedades da esperança matemática, verifica-se que o estimador da Equação (2.3) é não-viesado,

$$E[\widehat{\mu}] = E\left[\frac{1}{N}\sum_{i=1}^{N}\frac{Y_i I_i'}{\Pi_i'}\right] = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i E(I_i')}{\Pi_i'} = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i \Pi_i'}{\Pi_i'} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \mu \qquad (2.4)$$

O estimador da média $\overline{y}_{AD}$ para a $AAS$ adaptável é apresentado na Introdução deste trabalho. Esse é dado pela mudança no denominador da Equação (2.16) em relação a (1.12). Assim, tem-se que a função indicadora tem o mesmo papel de $f_i$ de (1.10) e de (2.2), e portanto verifica-se que esse estimador é viesado.

$$E[\overline{y}_{AD}] = E[\widehat{\mu}] = E\left[\frac{1}{n_1}\sum_{i=1}^{n_1} y_i\right] = \frac{E\left(\sum_{i=1}^{N} y_i I_i'\right)}{n_1} = \frac{\sum_{i=1}^{N} Y_i E(I_i')}{n_1} = \frac{\Pi_i' N\mu}{n_1} \neq \widehat{\mu} \qquad (2.5)$$

Para facilitar a análise da Equação (2.3) é mais conveniente reescrevê-la em função das distintas redes, pois a probabilidade de intersecção, $\Pi_i'$, é a mesma (também denominada $\alpha_k$) para cada unidade $i$ na $k$-ésima rede. Assim,

$$\alpha_k = 1 - \left[\binom{N - x_k}{n_1} \Big/ \binom{N}{n_1}\right] \qquad (2.6)$$

Analogamente à Equação (**??**) e como $p_{jk}$ é a probabilidade da $k$-ésima e $j$-ésima rede não se interseptarem, tem-se

$$p_{jk} \;=\; P(J_j \neq 1 \cap J_k \neq 1) = \binom{N - x_j - x_k}{n_1} \Big/ \binom{N}{n_1} \qquad (2.7)$$

Assim, das Equações (2.6) e (2.7), obtem-se $\alpha_{jk}$ como

$$\begin{aligned}
\alpha_{jk} &= \alpha_j + \alpha_k - (1 - p_{jk}) \\
&= 1 - \left[\binom{N - x_j}{n_1}\right] / \binom{N}{n_1} + 1 - \left[\binom{N - x_k}{n_1} / \binom{N}{n_1}\right] - \left[1 - \binom{N - x_j - x_k}{n_1} / \binom{N}{n_1}\right] \\
&= 1 - \frac{\binom{N-x_j}{n_1}}{\binom{N}{n_1}} + 1 - \frac{\binom{N-x_k}{n_1}}{\binom{N}{n_1}} - 1 + \frac{\binom{N-x_{j-x_k}}{n_1}}{\binom{N}{n_1}} \\
&= 1 - \left[\binom{N - x_j}{n_1} + \binom{N - x_k}{n_1} - \binom{N - x_j - x_k}{n_1}\right] / \binom{N}{n_1} \qquad (2.8)
\end{aligned}$$

Com isso,

$$\widehat{\mu} = \frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* J_k'}{\alpha_k} = \frac{1}{N} \sum_{k=1}^{K} \frac{y_k^*}{\alpha_k} \qquad (2.9)$$

onde $y_k^*$ é a soma dos valores de $y$ para a $k$-ésima rede, $K$ é o número total de redes distintas na população, $k$ é o número de redes distintas na amostra e $J_k$ assume o valor 1 com probabilidade $\alpha_k$ se a amostra inicial interseptar a $k$-ésima rede e 0 caso contrário.

Seja $z_k = y_k^*/\alpha_k$, $y_k^* = \sum_{i=1}^{N} y_i$ $\Pi_K = \alpha_k$ e $\Pi_{jk}$. Das Equações (1.3), (1.4), (1.5) e dessas definições, então pode-se obter o valor esperado e a variância da Equação (2.3),

$$E[\widehat{\mu}] = \frac{1}{N} \sum_{k=1}^{K} z_k E(J_k) = \frac{1}{N} \sum_{k=1}^{K} z_k \alpha_k = \frac{1}{N} \sum_{k=1}^{K} y_k^* = \frac{1}{N} \sum_{i=1}^{N} y_i^* = \overline{y} = \frac{\tau}{N} = \mu \qquad (2.10)$$

$$\begin{aligned}
var[\widehat{\mu}] &= var\left[\frac{1}{N} \sum_{k=1}^{K} z_k J_k\right] = \frac{1}{N^2}\left[\sum_{k=1}^{K} z_k J_k + \sum_{j=1}^{K} \sum_{j \neq k} cov\left(z_j J_j z_k J_k\right)\right] \\
&= \frac{1}{N^2}\left[\sum_{j=1}^{K} z_j^2 \Pi_j (1 - \Pi_k) + \sum_{j=1}^{K} \sum_{j \neq k} z_j z_k \Pi_{jk} - \Pi_i \Pi_k\right] \\
&= \frac{1}{N^2}\left[\sum_{j=1}^{K} \sum_{k=1}^{K} z_j z_k (\Pi_{jk} - \Pi_j \Pi_k)\right] \\
&= \frac{1}{N^2}\left[\sum_{j=1}^{K} \sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_j \alpha_k}\right)\right] \qquad (2.11)
\end{aligned}$$

bem como um estimador não-viesado da variância da Equação (2.11).

$$\begin{aligned}
\widehat{var}[\widehat{\mu}] &= \sum_{j=1}^{K}\sum_{k=1}^{K} z_j z_k J_j J_k \left(\frac{\Pi_{jk} - \Pi_j \Pi_k}{\Pi_{ij}}\right) \\
&= \frac{1}{N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_{jk}\alpha_j\alpha_k}\right) J_j J_k\right] \\
&= \frac{1}{N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk}}{\alpha_{jk}\alpha_j\alpha_k} - \frac{1}{\alpha_{jk}}\right)\right] \\
&= \frac{1}{N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} \frac{y_j^* y_k^*}{\alpha_{jk}} \left(\frac{\alpha_{jk}}{\alpha_j\alpha_k} - 1\right)\right] \qquad (2.12)
\end{aligned}$$

Outro estimador conhecido para a amostragem espacial por conglomerado adaptável é aquele que usa o número esperado de intersecção inicial e será abordado a seguir.

#### 2.3.1.2 Estimadores Usando o Número Esperado de Intersecção Inicial

O estimador dado pela Equação (2.3) pode ser reescrito na seguinte forma:

$$\widetilde{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i \frac{f_i}{E[f_i]} \qquad (2.13)$$

onde $f_i$ representa o número de unidades na amostra inicial que intercepta a rede $A_i$ que inclui a unidade $i$, $N$, o número de grades regulares. Se durante o processo de estimação fosse deixado de lado as unidades de borda dos conglomerados, $f_i$ seria interpretado como o número de vezes que a $i$-ésima unidade da amostra final aparece no estimador. Percebe-se, então, que $f_i = 0$ quando nenhuma unidade intercepta $A_i$ na amostra inicial.

O estimador na Equação (2.13) é não-viesado, pois

$$E\left[\widetilde{\mu}\right] = E\left[\frac{1}{N}\sum_{i=1}^{N} y_i \frac{f_i}{E[f_i]}\right] = \frac{1}{N}\sum_{i=1}^{N} E(y_i)\frac{E[f_i]}{E[f_i]} = \frac{1}{N}\sum_{i=1}^{N} y_i = \mu \qquad (2.14)$$

Como $m_i$ é o número de unidades na rede a que $i$ pertence e da Equação (1.77), pode ser encontrado outro estimador não-viesado que é definido por: como $f_i$ unidades é selecionado de $m_i$ unidades em $A_i$, então fica caracterizado que $f_i$ segue uma

distribuição Hipergeométrica com os seguintes parâmetros: $(N, m_i, n_1)$ (Thompson, 1991). Com isso, $E[f_i] = \frac{n_1 m_i}{N}$ e substituindo o valor dessa esperança na Equação (2.13) tem-se,

$$\widetilde{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i \frac{f_i}{\frac{n_1 m_i}{N}} = \frac{N}{N}\sum_{i=1}^{N} \frac{y_i f_i}{n_1 m_i} = \frac{1}{n_1}\sum_{i=1}^{N} \frac{y_i f_i}{m_i} \tag{2.15}$$

Para encontrar a variância do estimador da Equação (2.15) é utilizado a sua abordagem em termos de $n_1$ redes conectadas, não necessariamente distintas, pela amostra inicial. Como $m_i$ tem o mesmo valor para todas as unidades em $A_i$ e $w_i$ é a média das $m_i$ observações em $A_i$ (Thompson, 1990), então

$$\widetilde{\mu} = \frac{1}{n_1}\sum_{i=1}^{n_1} \frac{1}{m_i}\sum_{j \epsilon A_i} y_j = \frac{1}{n_1}\sum_{i=1}^{n_1} w_i = \overline{w} \tag{2.16}$$

Assim, $\widetilde{\mu}$ é a média amostral obtida por meio de uma seleção de uma $AAS$ de tamanho $n_1$ de uma população de $w_i$ valores. Dado que $w_i = \overline{v}_k$ é o mesmo para cada unidade na $k$-ésima rede, que há $x_k$ unidades nessa $k$-ésima rede e que $B_K$ é o conjunto de unidades na $k$-ésima rede, então

$$E(\widetilde{\mu}) = E(\overline{w}) = E\left[\frac{1}{N}\sum_{i=1}^{N} w_i\right] = \frac{1}{N}\sum_{k=1}^{K} x_k \overline{v}_k = \frac{1}{N}\sum_{k=1}^{K}\sum_{i \epsilon B_k} y_i = \mu \tag{2.17}$$

Das Equações (1.79) e (**??**), obtem-se a variância de (2.17) e um estimador não-viesado dessa variância

$$var[\widetilde{\mu}] = var\left[\frac{1}{N}\sum_{i=1}^{N} w_i\right] = \frac{N - n_1}{N n_1 (N-1)}\sum_{i=1}^{N}(w_i - \mu)^2 = \frac{\sigma^2}{n_1}\left(1 - \frac{n}{N}\right) \tag{2.18}$$

onde $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(w_i - \mu)^2$, $\Pi_{ij} = \frac{n_1(n_1-1)}{N(N-1)}$.

$$\widehat{var}[\widetilde{\mu}] = \frac{N - n_1}{N n_1 (n_1 - 1)}\sum_{i=1}^{n_1}(w_i - \widetilde{\mu})^2 \tag{2.19}$$

Como em algumas populações pode haver informações a priori sobre onde as agregações ocorrem, pode-se usar a técnica de amostragem estratificada para reduzir

a variância dos estimadores. Para isso será utilizada a teoria de amostragem espacial adaptável descrita acima com algumas modificações para a amostragem espacial adaptável estratificada por conglomerados, onde a população é dividida em estratos e o número de unidades são amostradas para cada estrato.

## 2.4 Amostragem Espacial Adaptável Estratificada por Conglomerados

No caso da amostragem espacial adaptável estratificada por conglomerados, também é preciso saber as coordenadas geográficas da área selecionada. Depois disso, estratifica-se a área usando as informações a priori e desenha a grade de toda a área selecionada (incluindo as áreas estratificadas) por meio das respectivas localizações, conforme indicado na Figura 2.3 (a). Após isso, seleciona-se uma amostra através da técnica de $AAS$, como na Figura 2.3 (b). Em seguida, identifica-se as unidades de interesse, como na 2.3 (c). Por último, agrega-se sucessivamente os vizinhos das áreas selecionadas: superiores, inferiores, da direita e da esquerda, até esgotar o critério de seleção da amostra adaptável, como na Figura 2.3 (d).

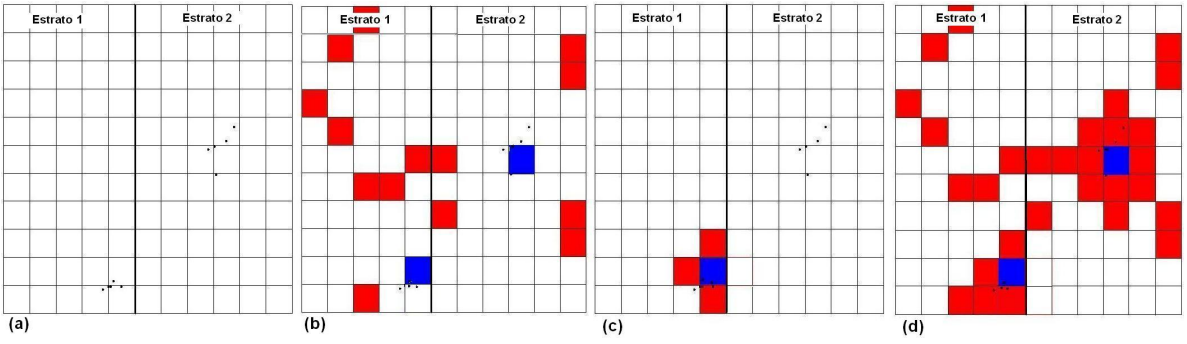


Figura 2.3: Passos da Amostragem Espacial Adaptável Estratificada por Conglomerados

Nesses estimadores a serem considerados, a amostra inicial também pode ser selecionada por meio de $AAS_C$ ou $AAS_S$. Na seção seguinte será abordado três estimadores não-viesados para esse tipo de amostragem.

### 2.4.1 Estimadores

Seja o total populacional de $N$ unidades particionado em $L$ estratos, com $n_h$ unidades no $h$-ésimo estrato, onde $(h = 1, 2, \ldots, L)$. Define-se a unidade $(h, i)$ como a $i$-ésima unidade no $h$-ésimo estrato associada ao valor de $y_{hi}$. Esse processo se inicia com a retirada de uma $AAS$ de $n_h$ unidades em cada estrato $L$, onde $n_0 = \sum_{h=1}^{L} n_h$ é o tamanho inicial da amostra. A partir disso, os conglomerados começam a ter os vizinhos adicionados de acordo com a condição $C$ estabelecida (Thompson e Seber, 1996).

#### 2.4.1.1 Estimadores Usando a Probabilidade de Intersecção Inicial

Usando o estimador baseado no probabilidade de interseção inicial (2.3), obtem-se

$$\widehat{\mu}_{st} = \frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* J_k}{\alpha_k} \tag{2.20}$$

onde as $K$ redes distintas têm legendas de $(1, 2, \ldots, k)$, $J_k$ é igual a 1 com probabilidade $\alpha_k$ se a amostra inicial de tamanho $n_0$ está na rede $k$ e 0 caso contrário e, por fim, $y_k$ é a soma dos valores de $y$ para a rede $k$.

Para definir $\alpha_k$ é necessário considerar as probabilidades de intersecção da rede $k$ com a amostra inicial em cada estrato. Além disso, define-se $x_{hk}$ como o número de unidades no estrato $h$ presente na rede $k$. Esse número assumirá o valor 0 se a rede $k$ aparecer fora do estrato $h$. Se a rede aparecer na borda de algum estrato, ignora-se essa unidade da rede. Assim, obtem-se

$$\alpha_k = 1 - \left[ \prod_{h=1}^{L} \frac{\binom{N_h - x_{hk}}{n_h}}{\binom{N_h}{n_h}} \right] \tag{2.21}$$

Para encontrar a variância do estimador da média não-viesado, define-se a probabilidade de a amostra inicial interceder a rede em $k$ e $k'$. Assim,

$$\alpha_{kk'} = 1 - (1 - \alpha_k) - (1 - \alpha_{k'}) + \left[ \prod_{h=1}^{L} \frac{\binom{N_h - x_{hk} - x_{hk'}}{n_h}}{\binom{N_h}{n_h}} \right] \tag{2.22}$$

Dado que $\alpha_{kk} = \alpha_k$ e da Equação (1.4), tem-se que

$$var[\widehat{\mu}_{st}] = \frac{1}{N^2} \sum_{k=1}^{K} \sum_{k'=1}^{K} y_k^* y_{k'}^* \left( \frac{\alpha_{kk'} - \alpha_k \alpha_{k'}}{\alpha_k \alpha_{k'}} \right) \tag{2.23}$$

e que um estimador não-viesado para essa variância é dado por

$$\widehat{var}[\widehat{\mu}_{st}] = \frac{1}{N^2} \sum_{k=1}^{K} \sum_{k'=1}^{K} y_k^* y_{k'}^* \left( \frac{\alpha_{kk'} - \alpha_k \alpha_{k'}}{\alpha_{kk'} \alpha_k \alpha_{k'}} \right) I_k I_{k'} \tag{2.24}$$

Outro estimador para esse tipo de amostragem é o que usa o número esperado de intersecção inicial, que será explicado na próxima seção.

#### 2.4.1.2 Estimadores Usando o Número Esperado de Intersecção Inicial

Seja $A_{hi}$, a rede que contém a unidade $(h, i)$, $u_{hi}$; $A_{ghi}$, a parte de $A_{hi}$ no estrato $g$; $f_{ghi}$, o número de unidades da amostra inicial no estrato $g$ que está em $A_{ghi}$; $m_{ghi}$, o número de unidades em $A_{ghi}$. Então, tem-se o número de unidades de uma amostra inicial de $n_0$ unidades que estão em $A_{hi}$ dado por (Thompson e Seber, 1996),

$$f_{.hi} = \sum_{g=1}^{L} f_{ghi} \tag{2.25}$$

Da Equação (2.13) obtem-se o estimador para a média

$$\widetilde{\mu}_{st} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} y_{hi} \frac{f_{.hi}}{E[f_{.hi}]} \tag{2.26}$$

Analogamente a Equação (2.4) e de (1.3), esse estimador é não-viesado.

Como o $f_i$ na Seção (2.3.1.2), aqui o $f_{ghi}$ tem uma distribuição hipergeométrica com parâmetros $(N_g, m_{ghi}, n_g)$. Logo, sabe-se que $E[f_{ghi}] = \frac{n_g m_{ghi}}{N_g}$ e $E[f_{f.hi}] = \sum_{i=1}^{L} \frac{n_g}{N_g} m_{ghi}$ (Thompson, 1991). Assim,

$$\widetilde{\mu}_{st} = \frac{1}{N} \sum_{h=1}^{L} y_{hi} \frac{f_{.hi}}{\sum_{i=1}^{L} \frac{n_g}{N_g} m_{ghi}} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \left( y_{hi} \sum_{g=1}^{L} f_{ghi} / \sum_{g=1}^{L} \frac{n_g}{N_g} m_{ghi} \right) \tag{2.27}$$

onde $f_{ghi}$ representa o número de unidades na amostra inicial que está na intersecção

do estrato $g$ com a unidade de rede a que a unidade $u_{hi}$ pertence.

Caso exista alguma coincidência de adicionar os mesmos vizinhos, obtem-se um estimador do estrato independente que combinado com o estimador com pesos fornece um estimador da média da população como o da segunda igualdade da Equação (1.42). Essa característica de agregação de vizinhos iguais gera uma perda da eficiência e um sistema mais eficiente seria permitir que os grupos sobreponham as fronteiras dos estratos (Thompson, 1991).

Assim, para encontrar a variância desse estimador da média, usa-se a Equação (2.16) para reescrever $\widetilde{\mu}_{st}$ em termos dos pesos das médias amostrais. Para isso, relacionam-se as observações ao intercepto das redes pela amostra inicial. Dessa forma, o termo $y_{hi}f_{.hi}$ significa que $A_{hi}$ é interceptado $f_{hi}$ vezes na amostra inicial, então $\widetilde{\mu}_{st}$ representa o soma ponderada de todas as unidades em todas as redes correspondentes à amostra inicial, com algumas redes sendo repetidas.

Seja $E[f_{.hi}]$ é o mesmo para cada unidade em $A_{hi}$, tem-se

$$\widetilde{\mu}_{st} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{n_h}\frac{1}{E[f_{.hi}]}\sum_{(h',i')\epsilon A_{hi}} y_{h'i'} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{n_h}\frac{Y_{hi}}{E[f_{.hi}]} \tag{2.28}$$

onde $Y_{hi}$ é a soma das $y$-ésima observações em $A_{hi}$.

Seja $\overline{w}_h = \sum_{i=1}^{n_h}\frac{w_{hi}}{n_h}$ e $w_{hi} = \frac{n_h Y_{hi}}{N_h E[f.hi]}$, então uma outra forma de reescrever Equação (2.28) é

$$\widetilde{\mu}_{st} = \sum_{h=1}^{L}\frac{N_h}{N}\overline{w}_h = \frac{1}{N}\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h} w_{hi} \tag{2.29}$$

onde $w_{hi} = \frac{Y_{hi}}{\sum_g m_{ghi}}$, quando $\frac{n_h}{N_h}$ tiver o mesmo valor para todos os estratos. Dessa forma, a Equação (2.29) representa uma média amostral estratificada de uma amostragem estratificada aleatória sem reposição, tendo como variável de interesse $w_{hi}$.

Então a variância desse estimador para a média é dado por

$$var[\widetilde{\mu}_{st}] = \frac{1}{N^2}\sum_{h=1}^{L} N_h(N_h - n_h)\frac{\sigma_h^2}{n_h} \tag{2.30}$$

onde $\sigma_h^2$ representa a variância populacional do estrato, ou seja,

$$\sigma_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (w_{hi} - \overline{W}_h)^2 \tag{2.31}$$

onde $\overline{W}_h = \frac{\sum_{i=1}^{n_h} w_{hi}}{n_h}$ é a média populacional do estrato.

Um estimador não-viesado da variância da média, (2.30), pode ser obtido por meio da substituição de $\sigma_h^2$ pela variância amostral, $s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (w_{hi} - \overline{w}_h)^2$.

#### 2.4.1.3 Estimadores que ignoram qualquer unidade no limite do estrato

Segundo Thompson (1991), o estimador que ignora a qualquer unidade adicionada que se situa no limite do estrato, é dado por:

$$\mu_{st}^{''} = \sum_{h=1}^{L} \frac{N_h}{N} \widetilde{\mu}_h = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \left( y_{hi} \sum_{g=1}^{L} \frac{N_g}{n_g} f_{ghi} / \sum_{g=1}^{L} m_{ghi} \right) \tag{2.32}$$

onde $\widetilde{\mu}_h = \sum_{i=1}^{n_h} \frac{w_{hi}^{''}}{n_h}$ e $w_{hi}^{''}$ é o total dos valores de $y$ na intersecção do estrato $h$ com $A_{hi}$ dividida pelo número de unidades na intersecção, ou seja, esse valor representa a média da rede para a parte da rede que está em $A_{hi}$ no estrato $h$.

A esperança de $\mu_{st}^{''}$ é dada por

$$E[\mu_{st}^{''}] = \sum_{h=1}^{L} \frac{N_h}{N} \mu_h = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} y_{hi} = \mu \tag{2.33}$$

A variância $var[\mu_{st}^{''}]$ é dada por

$$var[\widetilde{\mu}_{st}^{''}] = \frac{1}{N^2} \sum_{h=1}^{L} N_h (N_h - n_h) \frac{\sigma^2}{n_h} \tag{2.34}$$

onde a variância populacional do estrato é $\sigma_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (w_{hi}^{''} - \overline{W}_h)^2$ e a média populacional do estrato é $\overline{W}_h = \sum_{i=1}^{N_h} \frac{w_{hi}}{N_h}$.

O estimador não-viesado da variância $\widehat{var}[\mu_{st}^{''}]$ é dada pela expressão (2.29) substituindo $\sigma_h^2$ por $s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (w_{hi} - \overline{w}_h)^2$.

# Capítulo 3

# Algoritmo Computacional

## 3.1 Introdução

A implementação computacional da amostragem espacial adaptável requer pelo menos três passos: o desenvolvimento computacional do desenho de grades regulares, Figura 3.1 (a); seleção de áreas específicas da grade, isto é, identificar uma amostra e verificar em qual parte dessa grade estão os dados, Figura 3.1 (b); identificar a unidade, Figura 3.1 (c); identificar os vizinhos das áreas selecionadas: superiores, inferiores, da direita e da esquerda, Figura 3.1 (d).

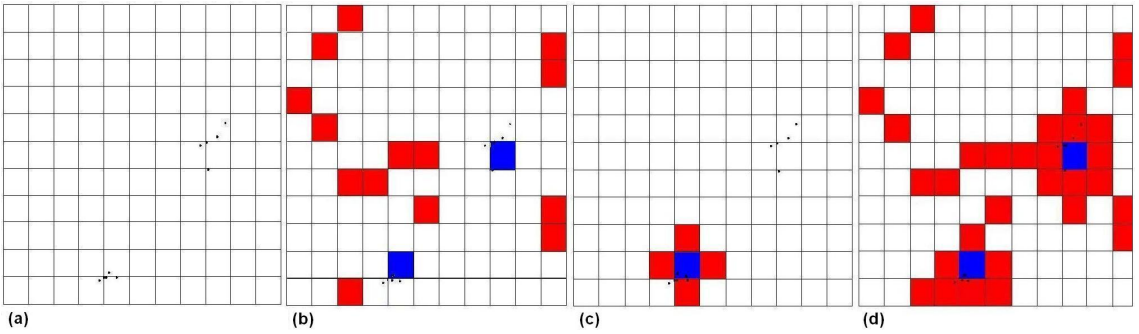


Figura 3.1: Principais Passos Computacionais da Amostragem Espacial Adaptável

Sendo assim, o trabalho visa desenvolver as macros no *software* SAS. Nesse Capítulo, apresenta-se o algoritmo computacional da amostragem espacial adaptável por conglomerado e da amostragem espacial adaptável estratificada por conglomerados no *software* SAS.

## 3.2 Desenho da Grade Regular

Para a criação da grade regular, no caso um quadrado, é necessário criar quatro pontos com as coordenadas inseridas no sentido horário (Pontos da Tabela 3.1) ou anti-horário.

Tabela 3.1: Tabela de Coordenadas do Quadrado

| **Referência** | **Valores** | ***Pontos*** |
|---|---|---|
| 1 | **(Min, Min)** | (0,0) |
| 2 | **(Min, Max)** | (0,1) |
| 3 | **(Max, Max)** | (1,1) |
| 4 | **(Max, Min)** | (1,0) |

No caso do quadrado, iniciou-se com os pontos no sentido horário, ou seja, a referência dos pontos na seguinte ordem:1, 2, 3, 4 gerou a Figura 3.2 (Quadrado). Caso essa ordem não seja seguida, o resultado é um polígono distorcido, conforme a Figura 3.2 (Polígono Distorcido).

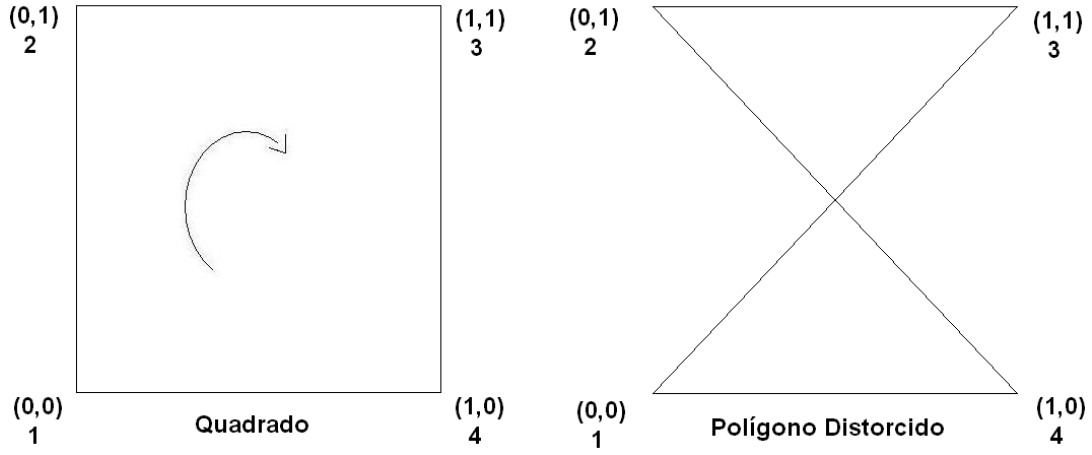


Figura 3.2: Quadrado e Polígono Distorcido

Como deseja-se desenhar uma grade em um área de estudo, é necessário conhecer os limites inferiores e superiores da região, ou seja, a coordenada mínima e máxima do eixo $y$ (latitude) e a coordenada mínima e máxima do eixo $x$ (longitude). A definição do tamanho de cada polígono é dada por:

```
%grid(minx=,maxx=,miny=,maxy=,dim=,anno=,printN=YES);
```

onde, os elementos da macro são: o valor mínimo da coordenada x, **MINX=**; o valor máximo da coordenada **x**, **MAXX=**. Analogamente, para as coordenadas **y** tem-se: **MINY=** e **MAXY=**. Outro parâmetro dessa macro é a dimensão do quadrado desenhado dado por **DIM=**. Por fim, os dois últimos elementos, **ANNO=** e **PRINTN=YES**, indicam a numeração de cada quadrado se o comando for **YES** (Figura 3.3 (a)) e **NO** (Figura 3.3 (b)) para não imprimir a numeração.

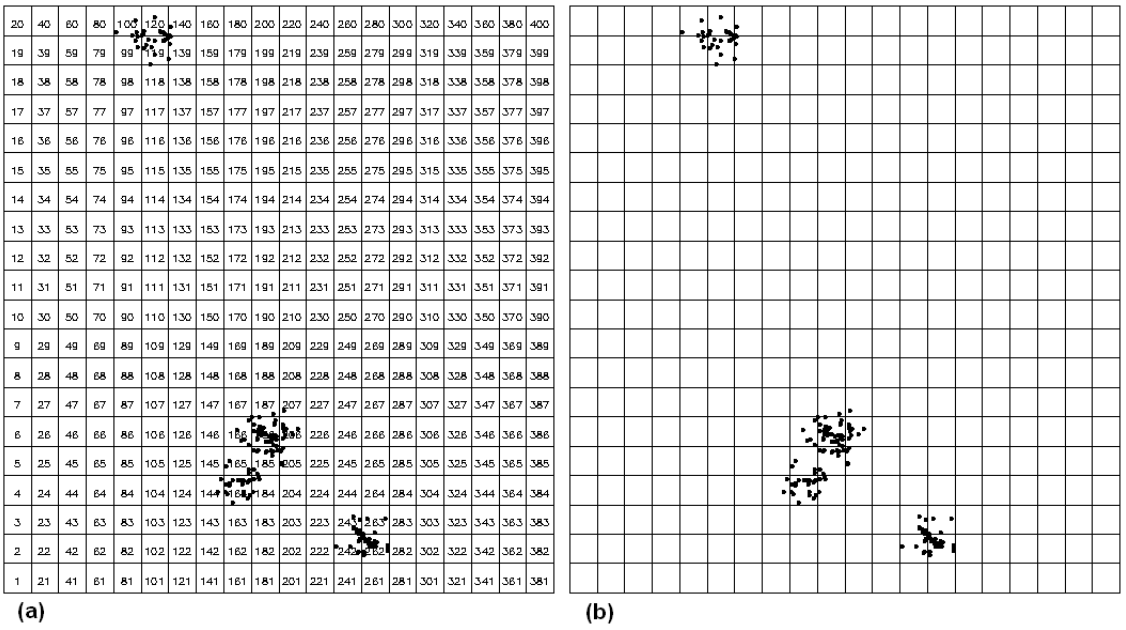


Figura 3.3: Grade Regular $N = 400$

A seguir, define-se o elemento, **id**, para as coordenadas do quadrado. Assim, o primeiro quadrado tem os pontos $(0,0), (0,1), (1,0), (1,1)$, **id=1** e assim sucessivamente. Isso é feito juntando a tabela com as coordenadas com a tabela criada a seguir:

```
data id&dim;
 do id=1 to &dim*&dim;
  do i=1 to 4;
   output;
  end;
 end;
run;
```

Essa numeração começa do valor numérico unitário e vai até o valor de $N$ a contar na direção vertical, iniciando da esquerda para a direita. No caso da Figura 3.3, verifica-se que a dimensão do quadrado é de $N = 20 \times 20 = 400$, variando o **id** de 1 a 400.

O passo seguinte é fazer a seleção de algumas áreas específicas da grade, ou seja, realiza-se uma *AAS* e caso essa área sorteada tenha unidades de interesse, seleciona-se os seus vizinhos.

## 3.3 Seleção de Áreas Específicas da Grade

A amostra na amostragem espacial adaptável corresponde aos polígonos da grade. Essa seleção pode ser feita por meio de um gerador de números aleatórios correspondente à quantidade de quadrados da grade. No SAS, isso pode ser feito pelo comando **PROC SURVEYSELECT** onde é obtida uma *AAS* de tamanho $n$ com um valor de semente dado pela variável **SEED**. Usa-se o comando **DATA** e **NOPRINT** para o *SAS* não imprimir esses valores, mas armazenar no banco de dados **DATA**.

```
proc surveyselect data=&data sampsize=&n out=&saida seed=&seed noprint;
run;
```

A identificação dos vizinhos das áreas selecionadas da próxima seção envolve três conceitos: verificação do ponto dentro do polígono; definição de vizinhos; e identificação dos polígonos vizinhos.

## 3.4 Identificação dos Vizinhos

A seleção de vizinhos é a parte computacional mais complicada e mais importante da técnica de amostragem espacial adaptável, uma vez que é a partir das áreas selecionadas que se inicia o processo de adaptação da amostra. Esse passo será descrito a seguir.

### 3.4.1 Verificação do Ponto dentro do Polígono

A verificação da existência de pontos dentro do polígono é mostrado na Figura 3.4. A ideia principal consiste em traçar uma semi-reta a partir do ponto $p$ de consulta até o infinito, em uma direção qualquer e verificar a quantidade de vezes que essa semi-reta passou pelas arestas desse polígono. Assim, se esse número de cruzamentos for ímpar, então o ponto está dentro do polígono (Kunigami, 2010).

A seleção desses pontos dentro da grade regular é feita pela macro **%ginside**.

```
%ginside(map=,id=,where=,data=,out=);
```

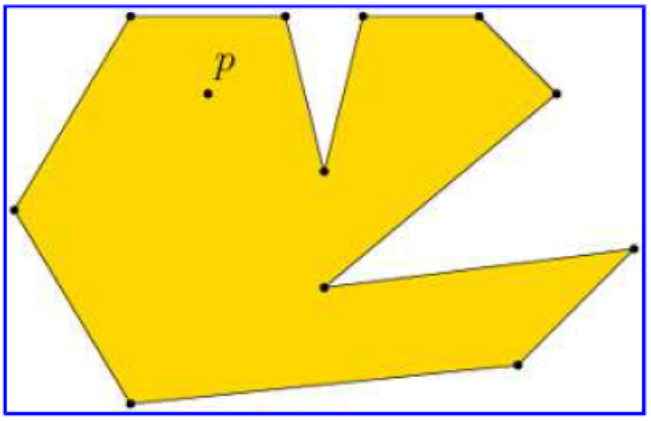


Figura 3.4: Ponto dentro do Polígono
Fonte:Kunigami (2010)

Essa macro foi feita da seguinte forma: Seja o polígono P formado por uma lista de pontos $\{p_1, p_2, \ldots, p_n\}$ em sentido horário. Com isso, um segmento é dado por um par de pontos $\{p_i, p_{i+1}\}$, incluindo o ponto $\{p_n, p_1\}$. Seja, ainda, um ponto $P$ que possui as coordenadas $(x, y)$. As condições ($if$ na sintaxe) estabelecem se a semi-reta do ponto $P$ intercepte algum segmento $(a, b)$, ou seja, de que esse ponto esteja verticalmente entre $a$ e $b$. Isso significa que $y_a \leq y_p \leq y_b$.

### 3.4.2 Definição de Vizinhos

A definição de vizinhos é abordada na Seção 2.2, ou seja, vizinhança é um conjunto de quadrados que serão adicionados na mesma amostra se esse quadrado da grade satisfizer a mesma condição de conter elementos de interesse no quadrado selecionado. Define-se vizinhança do tipo *ROOK* como sendo os polígonos que compartilham mais de um ponto em comum, no caso, os quadrados acima, abaixo, da direita e da esquerda, conforme a Figura 2.1.

A macro de vizinhança no *SAS* é dada por:

```
%neighborhood(id=,pt=,map=,anno=,out=,type=ROOK);
```

onde os seus elementos são: **ID=** os pontos das arestas dos quadrados; **PT=** a lista de pontos a ser formada; **MAP=** a tabela com os pontos extremos do quadrado; **ANNO=** a contagem dos quadrados; **OUT=** a tabela com todos os pontos vizinhos; **TYPE=ROOK**, indica que a forma padrão de seleção dos vizinhos é a do tipo **ROOK**,

mas tambem pode-se especificar o tipo **QUEEN**, que será explicada na Seção de contribuições para a amostragem espacial adaptável.

### 3.4.3 Identificação dos Polígonos Vizinhos

A identificação dos polígonos vizinhos é feita combinando a identificação dos vizinhos com os pontos dentro do polígono, formando a amostra final da amostragem espacial adaptável.

Por fim, tendo a base com as unidades selecionadas e suas respectivas contagens de pontos, o proximo passo é o cálculo dos estimadores apresentados no Capitulo 2.

## 3.5 Estimadores da Amostragem Espacial Adaptável

Nessa seção foram implementadas as fórmulas dos estimadores do Capítulo 2. Assim, para a amostragem espacial adaptável, implementou-se o estimador da média- **u1**, Equação 2.16; o estimador da variância do estimador da média- **varu1**, Equação 2.18; estimador do total- **Totu1**, dado pela multiplicação do total pelo estimador de **u1**, ou seja, **Totu1=NN**$\times$ **u1** e o estimador da variância de **Totu1**, denominada **TotVaru1**.

```
u1=(1/n)*(mu[,2]/mu[,3])‘*j(nrow(mu),1,1);
varu1=(NN-n)/(NN*n*(n-1))*((mu[,4]#((mu[,2]/mu[,3])-u1))‘*((mu[,2]/mu[,3])-u1));
Totu1=NN*u1;
TotVaru1=NN**2*varu1;
```

Analogamente, esses cálculos foram feitos para a *AAS*, estimador da média- **SRS** Equação 1.12; o estimador da variância do estimador da média- **VARSRS**, Equação 1.14; o estimador do total- **TotSRS**, Equação 1.15 e o o estimador da variância desse total **TotVarSRS**, Equação 1.19.

```
SRS=fsample[+,2]/n;
VARSRS=((1-n/NN)/n)*((fsample[,2]-SRS)‘*(fsample[,2]-SRS)/(n-1));
TotSRS=NN*SRS;
TotVarSRS=NN**2*varSRS;
```

Similarmente, foi calculado esses valores para a amostragem espacial adaptável com *AAS*, dado por:

```
adSRS=tab[+,2]/nrow(tab);
VARadSRS=((1-nrow(tab)/NN)/nrow(tab))*((tab[,2]-adSRS)‘*
(tab[,2]-adSRS)/(nrow(tab)-1));
TotadSRS=NN*adSRS;
TotVaradSRS=NN**2*VARadSRS;
```

Os estimadores da amostragem espacial adaptável estratificada por conglomerado foram calculados pela implementação das Equações da Seção 2.4.1.2: dos estimadores da média- 2.29, dos estimadores da variância dessa média- 2.30 e dos estimadores que ignoram qualquer unidade no limite do estrato da Seção 2.4.1.3 do Capítulo 2.

## 3.6 Contribuições para a Amostragem Espacial Adaptável

Nessa Seção será abordado uma outra forma de vizinhança que não foi considerada por (Thompson, 2002) na amostragem espacial adaptável: a vizinhança de matriz *QUEEN*, muito abordada em estudos de captura e recaptura.

### 3.6.1 Matriz *rook* e *queen*

Além da vizinhança com seleção do tipo *ROOK*, há outro tipo de vizinhaça: a do tipo *QUEEN*. Essa é definida como sendo os polígonos que compartilham pelo menos um ponto em comum, ou seja, nesse caso inclui-se os polígonos das diagonais, confrome a Figura 3.5.


Figura 3.5: Unidade Inicial com Seleção *Queen*

A Figura 3.5 é a Figura 2.1 incluindo as diagonais do quadrado. Isso significa que chegando o pesquisador na área a ser analisada, ele observa os pontos de interesse com uma visão de 360º.

Na amostragem espacial adaptável por conglomerados com seleção *QUEEN*, também é preciso saber as coordenadas geográficas da área selecionada. Depois disso, desenha a

grade de toda a área selecionada por meio das respectivas localizações, conforme indicado na Figura 3.6 (a). Após isso, seleciona-se uma amostra através da técnica de *AAS*, como na Figura 3.6 (b). Em seguida, identifica-se a unidade de interesse, como na Figura 3.6 (c). Por último, agrega-se sucessivamente os vizinhos das áreas selecionadas: superiores, inferiores, da direita, da esquerda e das diagonais, até esgotar o critério de seleção da amostra adaptável, como na Figura 3.6 (d).

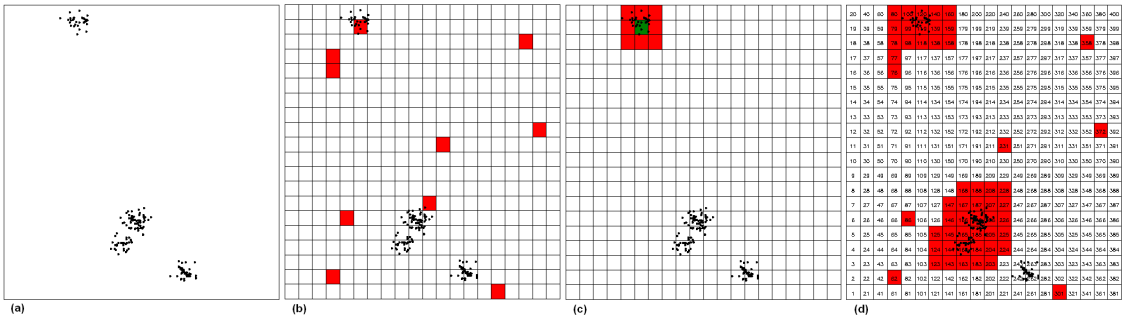


Figura 3.6: Passos da Amostragem Espacial por Conglomerados Adaptável com Seleção *QUEEN*

Com isso, a seleção da vizinhança na grade regular pode ser obtida usando-se a matriz *ROOK*, que é especificada na chamada da macro **%as** ou como *default*, dada por:

```
%as(data=a20,sample=amostra20,n=10,saida=adaps,seed=4,map=trab20,id=id,
anno=pontos2);
```

ou por meio da matriz *QUEEN*, especificando o tipo de seleção **TYPEN=QUEEN** na chamada macro **%as**.

```
%as(data=a20,sample=amostra20,n=10,saida=adaps,seed=4,map=trab20,id=id,
anno=pontos2,typen=queen);
```

Nesse caso, ressalta-se a existência de blocos na amostra final com matriz *QUEEN*, dado pela Figura 3.7 (a), diferentemente da amostra da *ROOK*, conforme a Figura 3.7 (b).

A macro completa desenvolvida no *software* SAS da amostragem espacial adaptável encontra-se no Apêndice A e os dados e as chamadas no Apêndice B.

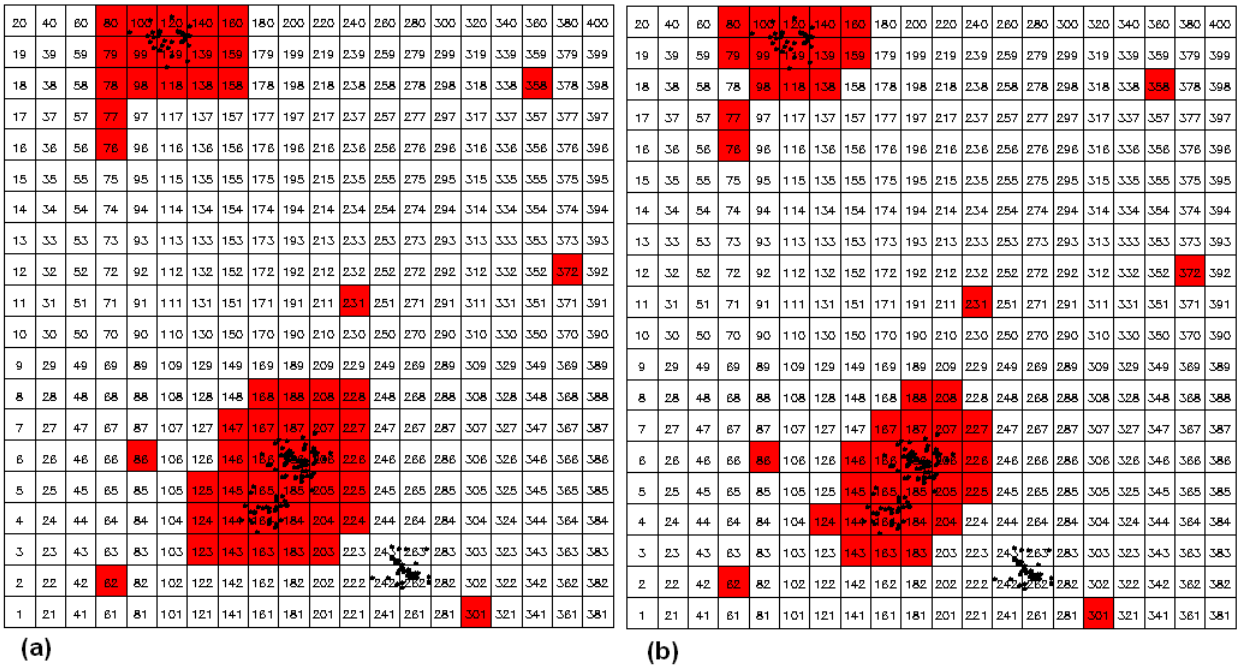


Figura 3.7: Amostra Final *QUEEN* (a) e *ROOK* (b)

# Capítulo 4

# Resultados

## 4.1 Introdução

Nesse Capítulo serão abordadas as análises obtidas por meio do estudo da técnica de amostragem espacial adaptável em simulações feitas no *software* SAS. Para isso, será implementado computacionalmente o exemplo da amostragem adaptável por conglomerado dado em Thompson (2002), a comparação entre os diferentes tamanhos da população ($N$ grades), bem como o exemplo da amostragem adaptável estratificada por conglomerados também dado em Thompson (2002).

## 4.2 Exemplo da Amostragem Adaptável por Conglomerado

Thompson (1990) apresenta um exemplo de como funciona a amostragem espacial adaptável e compara os resultados obtidos pela segunda igualdade da Equação (2.16), com o estimador da amostragem aleatória simples, $AAS$, dado por (1.12) e da $AAS$ com amostragem espacial adaptável, dado pela mudança no denominador da Equação (2.16) em relação a (1.12), conforme 2.5. Esse exemplo poderia representar uma reserva de animais que ficam agrupados (como manadas de elefantes) ou depósitos de minerais (como ouro, diamante, ferro) em grandes áreas.

Inicialmente é desenhada uma grade regular em cima da área a ser pesquisada e em seguida são selecionadas $n$ unidades (quadrados) pelo método AAS. Nesse exemplo, verifica-

se pela Figura 4.1(c) que a amostra inicial é formada por 10 unidades (total de quadrados na grade em vermelho), que foram selecionadas por meio de uma $AAS$ de um total de $N = 400$ unidades (que representam o número total de quadrados da grade regular onde cada lado tem tamanho igual a 20, ou seja, $20 \times 20 = 400$), com total de 190 pontos.

Selecionando os vizinhos das unidades iniciais que contém pelo menos uma unidade na amostra inicial, obtem-se a amostra final, Figura 4.1(d). A unidade superior tem um elemento, que intercepta a rede com $m_1 = 6$ unidades, que contém um total de $y_1^* = 36$ unidades de interesse. Outro ponto em que se verifica a unidade dentro do polígono que intercepta a rede com $m_2 = 11$ unidades, contém $y_2^* = 107$ unidades. Para as outras 8 unidades da amostra inicial, o valor de $y_i = 0$ e $m_i = 1$.

Há também 20 unidades de borda que não são utilizadas no cálculo das estimativas; aquelas selecionadas durante a seleção adaptável, mas que não contém unidades de interesse. Na Figura 4.1(d), as redes dentro dos dois grupos adicionados adaptativamente estão descritos na cor vermelha. As unidades restantes dentro de cada um desses grupos são unidades de borda.

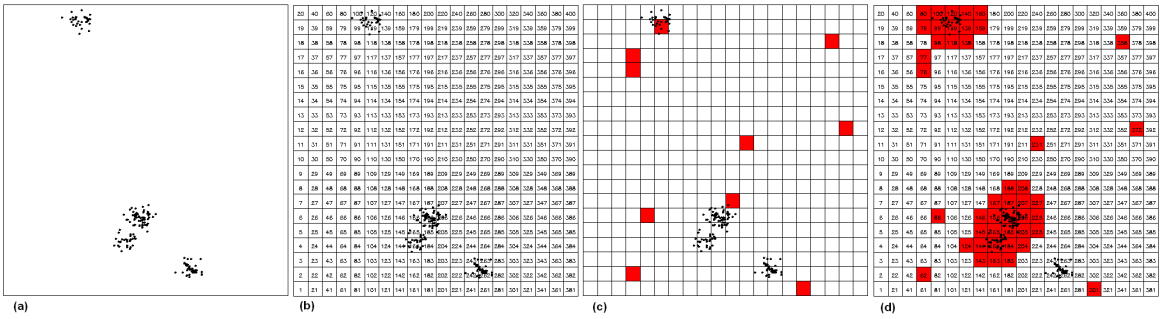


Figura 4.1: Exemplo da Amostragem por Conglomerados Adaptável

Para $w_1 = \frac{36}{6} = 6$ objetos por unidade, para $w_2 = \frac{107}{11} = 9,727$ e para os demais $w_i = 0$, calcula-se os valores para os estimadores da média, $\widetilde{\mu}$, e do total da amostragem adaptável:

$$\begin{aligned}
\widetilde{\mu} &= \frac{1}{10}\left[\frac{36}{6} + \frac{107}{11} + \left(\frac{0}{1}\right) + \ldots + \left(\frac{0}{1}\right)\right] = 1,573 \\
N\widetilde{\mu} &= 400 \times 1,573 = 629 \\
\widehat{var}[\widetilde{\mu}] &= \frac{(400-10)}{400(10)(10-1)}[(6-1,573)^2 + \ldots + (0-1,573)^2] = 1,147 \\
N^2\widehat{var}[\widetilde{\mu}] &= 400^2 \times 1,147 = 183.520
\end{aligned}$$

Para a $AAS$, onde $N$ são os números totais de quadrados da área selecionada e $n$ o número de quadrados selecionados da amostra inicial, obtém-se os valores para os estimadores da média $\overline{y}$ e do total $N\overline{y}$:

$$\overline{y} = \frac{11+1}{10} = 1,2$$
$$N\overline{y} = 400 \cdot 1,2 = 480$$
$$\widehat{var}[\overline{y}] = 1,165$$
$$N^2\widehat{var}[\overline{y}] = 186.506$$

Para as 45 unidades, que incluem as 25 unidades de borda da amostra final, calcula-se os valores para os estimadores da média, $\overline{y}_{AD}$ para a $AAS$ com adaptável, ou seja, usa a quantidade da Equação (1.12) em (2.16). Assim,

$$\overline{y}_{AD} = \frac{143}{45} = 3,178$$
$$N\overline{y}_{AD} = 400 \times 3,178 = 1.271$$
$$\widehat{var}[\overline{y}_{AD}] = 1,004$$
$$N^2\widehat{var}[\overline{y}_{AD}] = 400^2 \cdot 1,004 = 160.687$$

A Tabela 4.1 apresenta as estimativas encontradas e verifica-se que a variância, da média e do total, da amostragem $AAS$ adaptável é a menor entre as demais. Entretanto, a sua estimativa da média é a maior, pois há um viés ao usar o estimador da $AAS$ nessa amostragem. Comparando a razão entre as variâncias $\overline{y}_{AD}$ e $\widetilde{\mu}$, observa-se que há uma redução de 13% desta. Assim, considerando que se tem um total de 190 pontos e que a média real da população é de $\mu = \frac{190}{400} = 0,475$, a amostragem espacial adaptável nesse caso ficou bem próxima da AAS, mas como será visto mais adiante, a amostragem espacial adaptável varia muito menos quando $N$ varia.

Tabela 4.1: Tabela de Comparação dos Estimadores da Amostragem Adaptável, $AAS$, $AAS$ Adaptável

| **Estimador** | $\widetilde{\mu}$ | $\overline{y}$ | $\overline{y}_{AD}$ |
|---|---|---|---|
| **Média** | 1,57 | 1,20 | 3,17 |
| **Total** | 629 | 480 | 1.271 |
| **Variância da Média** | 1,147 | 1,165 | 1,004 |
| **Variância do Total** | 183.520 | 186.506 | 160.687 |

A saída computacional do programa implementado no *software* SAS para esse exemplo é dado pela Figura 4.2. Assim, verifica-se o número de observações ($n = 10$), o tamanho da população ($N = 400$), as estatísticas para os estimadores da média e do total, bem como os estimadores das suas respectivas variâncias nos três casos analisados: amostragem espacial adaptável por conglomerado, $AAS$, amostragem $AAS$ adaptável (viesada).

```
                Adaptive Cluster Sampling

          Number of Observations:          10

            Population Size:          400

                Adaptive Cluster Sampling

                        Statistics

        Mean Var of Mean         Sum Var of Sum

1.5727273   1.1470888 629.09091   183534.21

                Simple Random Sampling

                        Statistics

        Mean Var of Mean         Sum Var of Sum

     1.2    1.1656667         480   186506.67

        Adaptive Simple Random Sampling (biased)

                        Statistics

        Mean Var of Mean         Sum Var of Sum

3.1777778   1.0042994 1271.1111    160687.9
```

Figura 4.2: Saída do Exemplo da Amostragem Adaptável por Conglomerado

A próxima Seção irá mostrar a influência da quantidade de grades nos valores das estimativas.

## 4.3 Comparação entre os Diferentes Tamanhos da População

Nessa Seção será verificado se há interferência da variação do total de áreas (grades) nas estimativas, ou seja, quando $N$ variar. Assim, simulou-se diversos tamanhos de quadrados, e as amostras inicias foram definidas a fim de se ter a mesma amostra do exemplo de (Thompson, 1990). Dessa forma, obteve-se a Tabela 4.2 com os respectivos valores dos estimadores da média e de suas variâncias estimadas.

Tabela 4.2: Tabela de Comparação dos Estimadores da Média da Amostragem Adaptável, $AAS$, $AAS$ Adaptável com a Variação da População ($ROOK$)

| **Matriz** | $N$ | $\overline{\mu}$ | $\widehat{var}(\overline{\mu})$ | $\overline{y}$ | $\widehat{var}(\overline{y})$ | $\overline{y}_{AD}$ | $\widehat{var}(\overline{y}_{AD})$ |
|---|---|---|---|---|---|---|---|
| **4x4** | 16 | 2,82 | 1,35 | 7,60 | 10,30 | 11,87 | 0 |
| **5x5** | 25 | 3,08 | 3,43 | 11,60 | 42,33 | 10,00 | 6,06 |
| **6x6** | 36 | 3,48 | 5,99 | 12,60 | 62,66 | 8,26 | 6,64 |
| **7x7** | 49 | 5,36 | 11,57 | 3,30 | 3,89 | 5,96 | 5,50 |
| **8x8** | 64 | 3,40 | 4,70 | 4,00 | 7,52 | 6,78 | 4,28 |
| **9x9** | 81 | 3,72 | 5,41 | 10,40 | 47,76 | 5,93 | 3,88 |
| **10x10** | 100 | 3,58 | 5,13 | 0,80 | 0,27 | 4,76 | 4,01 |
| **11x11** | 121 | 2,72 | 3,01 | 5,40 | 13,54 | 4,61 | 2,50 |
| **12x12** | 144 | 2,98 | 3,85 | 1,80 | 1,42 | 4,61 | 3,13 |
| **13x13** | 169 | 2,73 | 3,10 | 9,00 | 36,88 | 4,76 | 4,51 |
| **14x14** | 196 | 2,13 | 2,31 | 1,90 | 1,95 | 4,08 | 1,83 |
| **15x15** | 225 | 2,73 | 3,22 | 4,80 | 10,82 | 4,20 | 1,77 |
| **16x16** | 256 | 2,09 | 1,91 | 6,90 | 22,69 | 3,67 | 1,71 |
| **17x17** | 289 | 1,79 | 1,44 | 5,00 | 15,55 | 3,49 | 1,42 |
| **18x18** | 324 | 1,79 | 1,45 | 1,30 | 0,73 | 3,40 | 1,06 |
| **19x19** | 361 | 2,06 | 2,03 | 4,00 | 7,69 | 3,76 | 1,21 |
| **20x20** | 400 | 1,57 | 1,15 | 1,20 | 1,16 | 3,18 | 1,00 |

A Figura 4.3 representa como a estimativa da média é influenciada nos casos da variação da população. Assim, a amostragem espacial adaptável por conglomerados (linha verde sólida) apresentou a menor interferência na média com a variação da população, tendo um decrescimento ao aumentar o tamanho da grade regular, salvo quando $N = 7$. A $AAS$ (linha azul pontilhada) teve uma variação durante todo o processo. Já a amostragem adaptável com a $AAS$ -$AASAD$ (linha vermelha tracejada) indicou um decrescimento com um leve crescimento para $N = 8$.

A Figura 4.4 representa como a estimativa da variância da média é influenciada nos casos da variação da população. Dessa forma, verifica-se que a variância da amostragem

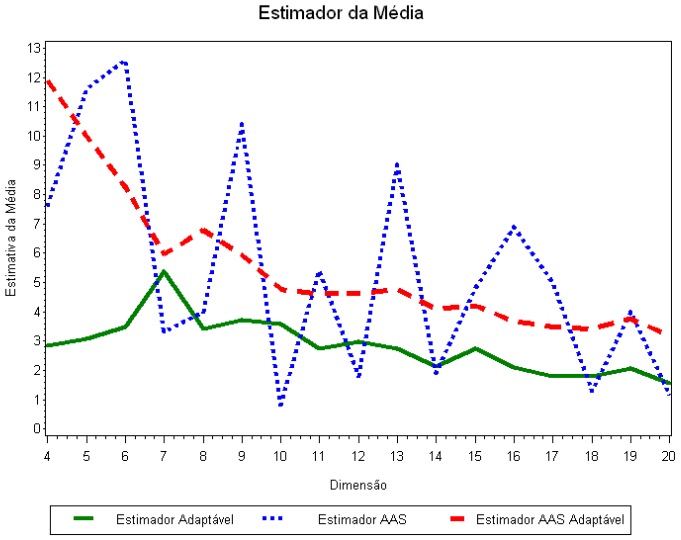


Figura 4.3: Análise da Média com a Variação da População ($ROOK$)

adaptável é a que apresenta uma variação de 0 a 15, enquanto a $AAS$ varia entre 10 a 75 e a $AASAD$ entre 0 a 15.

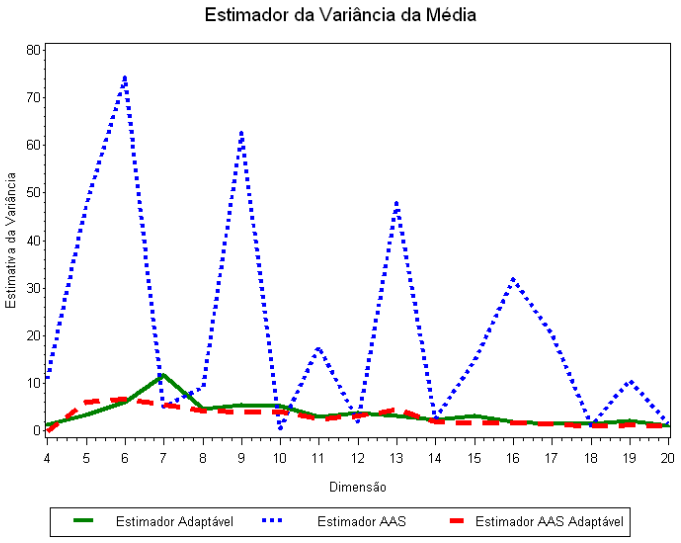


Figura 4.4: Análise da Variância da Média com a Variação da População ($ROOK$)

Analogamente, obtem-se a Tabela 4.3 para os estimadores do total. A Figura 4.5 representa como o estimador do total se comporta com a variação da população. Para esse caso, o estimador do total da amostragem adaptável, $N\mu$, é o que apresenta uma menor variação ao ser comparado com os outros dois ($N\overline{y}$ e $N\overline{y}_{AD}$). O estimador $N\overline{y}$ teve uma

variação durante todo o crescimento da população.

Tabela 4.3: Tabela de Comparação dos Estimadores do Total da Amostragem Adaptável, *AAS*, *AAS* Adaptável com a Variação da População (*ROOK*)

| **Matriz** | $N$ | $N\overline{\mu}$ | $N^2\widehat{var}(\overline{\mu})$ | $N\overline{y}$ | $N^2\widehat{var}(\overline{y})$ | $N\overline{y}_{AD}$ | $N^2\widehat{var}(\overline{y}_{AD})$ |
|---|---|---|---|---|---|---|---|
| **4x4** | 16 | 45,20 | 345,46 | 121,60 | 2.637,23 | 190,00 | 0 |
| **5x5** | 25 | 77,02 | 2.143,20 | 290,00 | 26.460,00 | 250,00 | 3.786,84 |
| **6x6** | 36 | 125,28 | 7.765,19 | 453,60 | 81.207,36 | 297,40 | 8.605,68 |
| **7x7** | 49 | 262,97 | 27.775,32 | 161,70 | 9.344,79 | 291,96 | 13.224,14 |
| **8x8** | 64 | 217,60 | 19.232,25 | 256,00 | 30.796,80 | 434,29 | 17.532,08 |
| **9x9** | 81 | 301,72 | 35.516,02 | 842,40 | 313.391,16 | 480,94 | 25.485,92 |
| **10x10** | 100 | 358,33 | 51.362,50 | 80,00 | 2.760,00 | 476,00 | 40.096,05 |
| **11x11** | 121 | 330,16 | 44.136,80 | 653,40 | 198.241,56 | 558,16 | 36.596,89 |
| **12x12** | 144 | 429,60 | 79.976,56 | 259,20 | 29.501,44 | 664,25 | 64.971,73 |
| **13x13** | 169 | 461,13 | 88.303,16 | 1.521,00 | 1.053.343,20 | 805,57 | 128.886,39 |
| **14x14** | 196 | 417,20 | 89.039,35 | 372,40 | 74.896,83 | 800,80 | 70.609,67 |
| **15x15** | 225 | 613,93 | 162.971,10 | 1.080,00 | 548.035,00 | 946,32 | 89.533,50 |
| **16x16** | 256 | 534,75 | 125.050,12 | 1.766,40 | 1.486.863,40 | 938,67 | 111.878,68 |
| **17x17** | 289 | 517,31 | 120.309,52 | 1.445,00 | 1.299.055,00 | 1.007,97 | 118.269,46 |
| **18x18** | 324 | 579,96 | 151.800,29 | 421,20 | 76.980,24 | 1.103,14 | 111.112,35 |
| **19x19** | 361 | 742,75 | 265.244,93 | 1.444,00 | 1.002.424,80 | 1.358,50 | 157.809,51 |
| **20x20** | 400 | 629,09 | 183.534,21 | 480,00 | 186.506,67 | 1.271,11 | 160.687,90 |

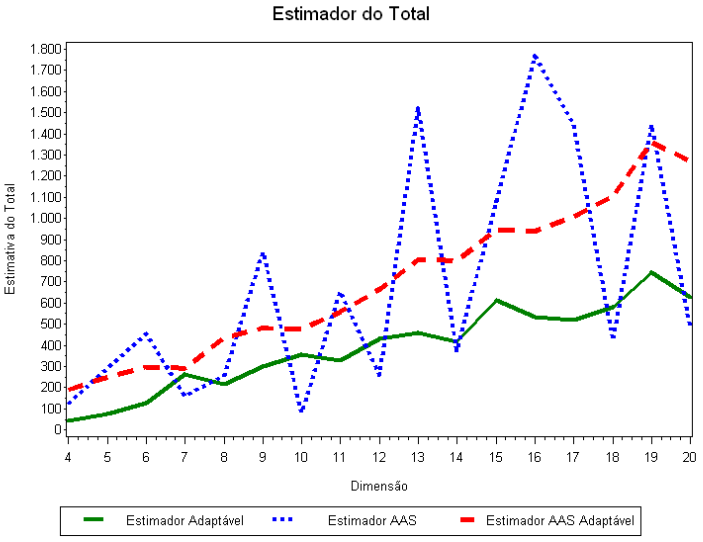


Figura 4.5: Análise do Total com a Variação da População (*ROOK*)

A Figura 4.6 representa como o estimador da variância do estimador do total se altera com a variação da população. Assim, o estimador $N^2\widehat{var}(\overline{\mu})$ e o $N^2\widehat{var}(\overline{y}_{AD})$ são bem próximos e o $N^2\widehat{var}(\overline{y})$ teve uma grande variação durante todo o processo.

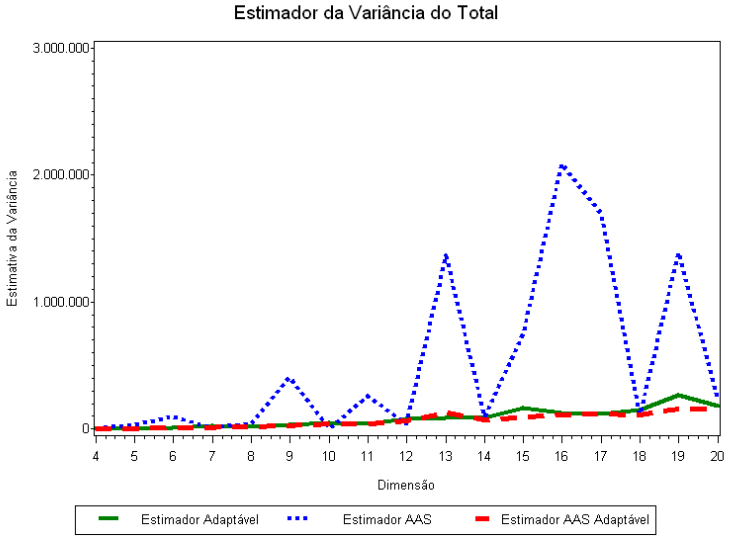


Figura 4.6: Análise da Variância do Total com a Variação da População ($ROOK$)

A Figura 4.7 representa como o estimador da variância do estimador do total se altera com a variação da população, retirando os grandes valores do $N^2\widehat{var}(\overline{y})$. Verifica-se que essa variância não é constante como no gráfico da Figura 4.6, apresentando um crescimento desordenado para a $AAS$ e um crescimento parecido para os outros dois casos.

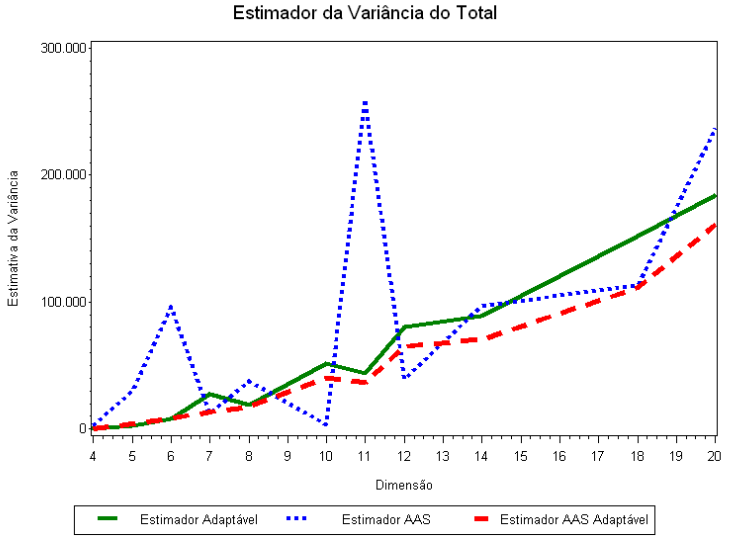


Figura 4.7: Análise da Variância do Total com a Variação da População ($ROOK$)

Na seção a seguir serão abordado os resultados para as diferentes formas de seleção de uma vizinhança, matriz $ROOK$ e $QUEEN$.

## 4.4 Matriz *ROOK* e *QUEEN*

Pelo exemplo da amostragem adaptável por conglomerado, obteve-se as Tabelas 4.4 para o caso de $n = 10, N = 20 \times 20$ com seleção *ROOK* e a Tabela 4.5 para a seleção *QUEEN*. Comparando essas Tabelas, verifica-se que os valores das estimativas de $\widetilde{\mu}$ são bem próximos tanto para a seleção *ROOK* quanto para a *QUEEN*.

Tabela 4.4: Tabela de Comparação dos Estimadores da Amostragem Espacial Adaptável por Conglomerados, *AAS*, *AAS* Adaptável com seleção *ROOK*

| **Estimador** | $\widetilde{\mu}$ | $\overline{y}$ | $\overline{y}_{AD}$ |
|---|---|---|---|
| **Média** | 1,57 | 1,20 | 3,17 |
| **Total** | 629 | 480 | 1.271 |
| **Variância da Média** | 1,147 | 1,165 | 1,004 |
| **Variância do Total** | 183.534 | 186.506 | 160.687 |

Dessa forma, constata-se que o estimador $\widetilde{\mu}$ é robusto para a amostragem espacial adaptável por conglomerado. Isso já era esperado, pois esse estimador é ponderado pela quantidade de quadrados $N$.

Tabela 4.5: Tabela de Comparação dos Estimadores da Amostragem Espacial Adaptável por Conglomerados, *AAS*, *AAS* Adaptável com seleção *QUEEN*

| **Estimadores/Valores** | $\widetilde{\mu}$ | $\overline{y}$ | $\overline{y}_{AD}$ |
|---|---|---|---|
| **Média** | 1,57 | 1,20 | 2,65 |
| **Total** | 629 | 480 | 1.059 |
| **Variância da Média** | 1,147 | 1,165 | 0,700 |
| **Variância do Total** | 183.534 | 186.506 | 112.011 |

Assim, não é necessário analisar as áreas situadas nas diagonais da unidade de interesse. Porém, verifica-se para os dois tipos de seleção que há um viés no estimador $\overline{y}_{AD}$. Isso se deve ao fato do seu cálculo considerar tanto áreas com grande quantidade de elementos de interesse quanto outras áreas que tem poucos elementos e outros que não têm.

### 4.4.1 Comparação entre os Diferentes Tamanhos da População com Seleção QUEEN

Analogamente à Seção 4.3, nessa seção será verificado a interferência da variância ao ser selecionado um menor ou maior tamanho de área para a seleção *QUEEN*. Dessa forma, obteve-se a Tabela 4.6 com os respectivos valores dos estimadores da média e de suas variâncias estimadas.

Tabela 4.6: Tabela de Comparação dos Estimadores da Amostragem Adaptável, $AAS$, $AAS$ Adaptável ($QUEEN$)

| **Matriz** | $N$ | $\overline{\mu}$ | $\widehat{var}(\overline{\mu})$ | $\overline{y}$ | $\widehat{var}(\overline{y})$ | $\overline{y}_{AD}$ | $\widehat{var}(\overline{y}_{AD})$ |
|---|---|---|---|---|---|---|---|
| **4x4** | 16 | 2,74 | 1,06 | 7,60 | 10,30 | 11,87 | 0 |
| **5x5** | 25 | 2,61 | 1,67 | 11,60 | 42,33 | 8,64 | 2,31 |
| **6x6** | 36 | 3,17 | 4,29 | 12,60 | 62,66 | 7,60 | 4,80 |
| **7x7** | 49 | 3,77 | 5,47 | 3,30 | 3,89 | 6,33 | 3,48 |
| **8x8** | 64 | 3,40 | 4,70 | 4,00 | 7,52 | 5,58 | 2,50 |
| **9x9** | 81 | 2,91 | 3,26 | 10,40 | 47,76 | 5,28 | 2,86 |
| **10x10** | 100 | 3,72 | 5,56 | 0,80 | 0,27 | 4,52 | 2,01 |
| **11x11** | 121 | 2,74 | 3,04 | 5,40 | 13,54 | 4,32 | 1,37 |
| **12x12** | 144 | 2,60 | 2,75 | 1,80 | 1,42 | 4,04 | 1,41 |
| **13x13** | 169 | 2,38 | 2,26 | 9,00 | 36,88 | 3,87 | 1,87 |
| **14x14** | 196 | 2,04 | 2,21 | 1,90 | 1,95 | 3,32 | 1,20 |
| **15x15** | 225 | 2,42 | 2,65 | 4,80 | 10,82 | 3,32 | 1,10 |
| **16x16** | 256 | 1,90 | 1,63 | 6,90 | 22,69 | 3,11 | 1,21 |
| **17x17** | 289 | 1,79 | 1,44 | 5,00 | 15,55 | 2,75 | 0,87 |
| **18x18** | 324 | 1,79 | 1,45 | 1,30 | 0,73 | 2,86 | 0,75 |
| **19x19** | 361 | 2,06 | 2,03 | 4,00 | 7,69 | 2,98 | 0,77 |
| **20x20** | 400 | 1,57 | 1,15 | 1,20 | 1,16 | 2,65 | 0,70 |

A Figura 4.8 representa como a estimativa da média é influenciada nos casos da variação da população. Assim, a amostragem espacial adaptável por conglomerados apresentou a menor interferência na média com a variação da população, tendo um decrescimento ao aumentar o tamanho da grade regular, salvo quando $N = 15$ e $n = 19$. A $AAS$ teve uma oscilação durante todo o processo, principalmente quando $N = 13$, $N = 16$ e $N = 19$. Já a amostragem adaptável com a $AAS$ -$AASAD$ indicou um decrescimento constante até $N = 12$.

A Figura 4.9 representa como a estimativa da variância da média é influenciada nos casos da variação da população. Assim, essa Figura indica que a variância da amostragem adaptável é a que apresenta uma variação de 0 a 5, enquanto a $AAS$ varia entre 10 a 60 e a $AASAD$ entre 0 a 5.

Similarmente, obtem-se a Tabela 4.7 para os estimadores do total para a amostragem espacial adaptável. A Figura 4.10 representa como o estimador do total se comporta com a variação da população. Para esse caso, o estimador do total da amostragem adaptável é o que apresenta uma menor variação ao ser comparado com os outros dois estimadores do total da $AAS$ e do estimador do total adaptável com $AAS$. O estimados do total da $AAS$ teve uma variação durante todo o crescimento da população enquanto ocrescimento

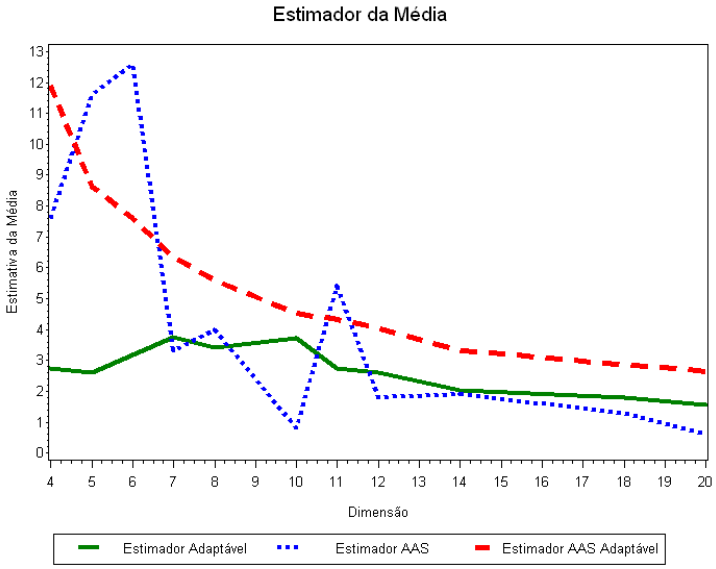


Figura 4.8: Análise da Média com a Variação da População ($QUEEN$)

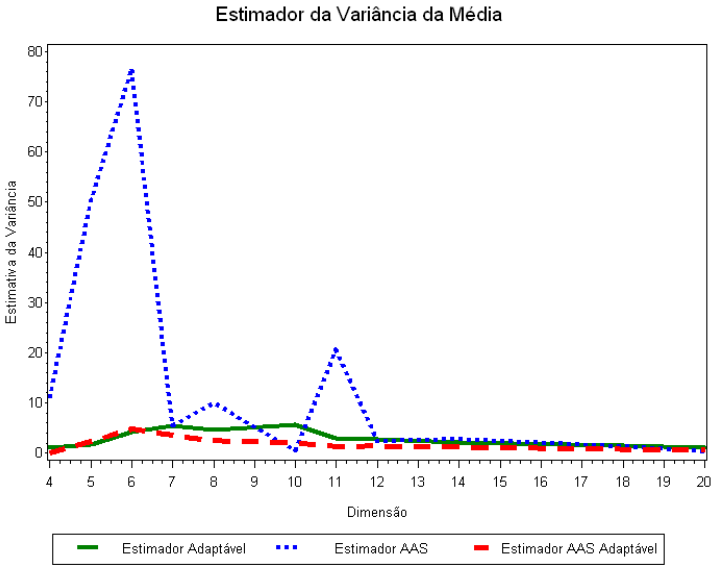


Figura 4.9: Análise da Variância da Média com a Variação da População ($QUEEN$)

do do estimador do total adaptável com $AAS$ foi constante até $N = 12$.

A Figura 4.11 representa como o estimador da variância do estimador do total se altera com o aumento da população. Assim, o estimador da variância do total para a amostragem espacial adaptável e o da amostragem espacial adaptável com $AAS$ são bem próximos e constantes até $N = 12$. Já o estimador da variância do total para $AAS$ teve uma grande

Tabela 4.7: Tabela de Comparação dos Estimadores do Total da Amostragem Adaptável, *AAS*, *AAS* Adaptável com a Variação da População (*QUEEN*)

| **Matriz** | $N$ | $N\overline{\mu}$ | $N^2\widehat{var}(\overline{\mu})$ | $N\overline{y}$ | $N^2\widehat{var}(\overline{y})$ | $N\overline{y}_{AD}$ | $N^2\widehat{var}(\overline{y}_{AD})$ |
|---|---|---|---|---|---|---|---|
| **4x4** | 16 | 43,84 | 270,43 | 121,60 | 2.637,23 | 190,00 | 0 |
| **5x5** | 25 | 65,28 | 1.046,12 | 290,00 | 26.460,00 | 215,91 | 1.443.69 |
| **6x6** | 36 | 114,00 | 5.557,07 | 453,60 | 81.207,36 | 273,60 | 6.223,80 |
| **7x7** | 49 | 184,57 | 13.129,35 | 161,70 | 9.344,79 | 310,33 | 8.364,73 |
| **8x8** | 64 | 217,60 | 19.232,25 | 256,00 | 30.796,80 | 357,65 | 10.233,56 |
| **9x9** | 81 | 235,80 | 21.412,59 | 842,40 | 313.391,16 | 427,50 | 18.789,75 |
| **10x10** | 100 | 372,50 | 55.580,62 | 80,00 | 2.760,00 | 452,38 | 20.109,62 |
| **11x11** | 121 | 331,54 | 44.557,55 | 653,40 | 198.241,56 | 522,50 | 20.020,45 |
| **12x12** | 144 | 374,40 | 56.953,22 | 259,20 | 29.501,44 | 582,13 | 29.240,51 |
| **13x13** | 169 | 403,00 | 64.518,67 | 1.521,00 | 1.053.343,20 | 655,31 | 53.366,62 |
| **14x14** | 196 | 400,40 | 85.073,26 | 372,40 | 74.896,80 | 651,81 | 46.018,82 |
| **15x15** | 225 | 546,43 | 134.254,34 | 1.080,00 | 548.035,00 | 748,26 | 55.857,11 |
| **16x16** | 256 | 488,67 | 107.131,45 | 1.766,40 | 1.486.863,40 | 795,83 | 79.587,04 |
| **17x17** | 289 | 517,31 | 120.309,52 | 1.445,00 | 1.299.055,00 | 794,75 | 72.618,86 |
| **18x18** | 324 | 579,96 | 151.800,29 | 421,20 | 76.980,24 | 926,64 | 78.703,36 |
| **19x19** | 361 | 742,75 | 265.244,93 | 1.444 | 1.002.424,80 | 1.075,48 | 100.921,34 |
| **20x20** | 400 | 629,09 | 183.534,21 | 480,00 | 186.506,67 | 1.059,26 | 112.011,87 |

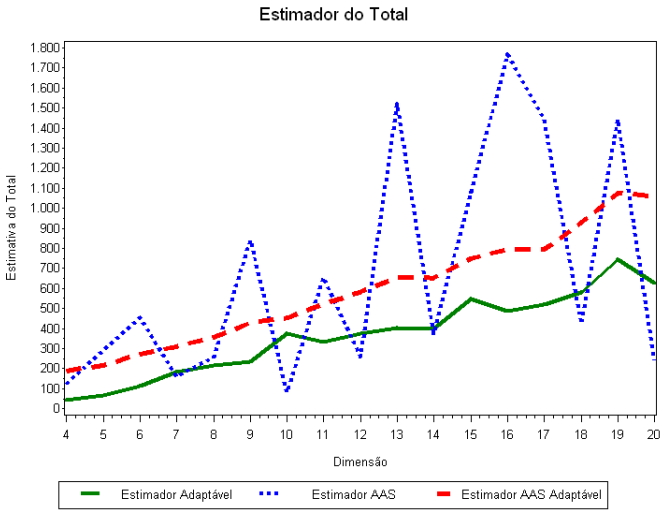


Figura 4.10: Análise do Total com a Variação da População (*QUEEN*)

variação durante todo o processo, principalmente quando $N = 13$, $N = 16$ e $N = 19$.

A Figura 4.12 representa como o estimador da variância do estimador do total se altera com o aumento da população, retirando os grandes valores do $N^2\widehat{var}(\overline{y})$. Verifica-se que essa variância não é constante como no gráfico da Figura 4.11, apresentando um

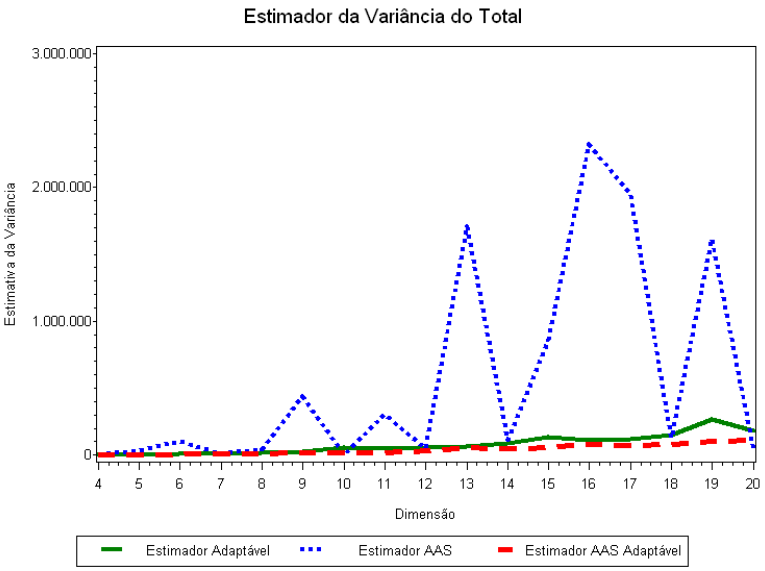


Figura 4.11: Análise da Variância do Total com a Variação da População ($QUEEN$)

crescimento desordenado para a $AAS$ e um crescimento parecido para os outros dois casos.

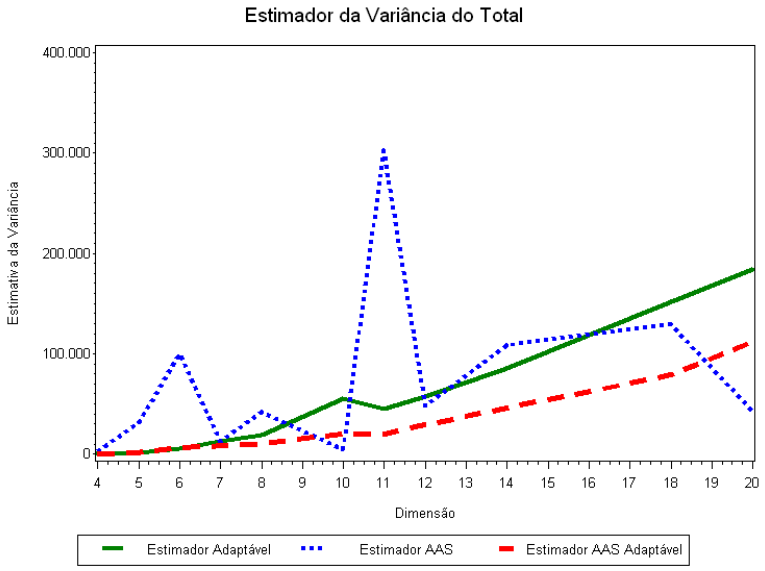


Figura 4.12: Análise da Variância do Total com a Variação da População ($QUEEN$)

Por fim, As Figuras 4.13 e 4.14 representam como o estimador da média e do total, tanto com seleção $ROOK$ quanto $QUEEN$, se comportam com a variação da população.

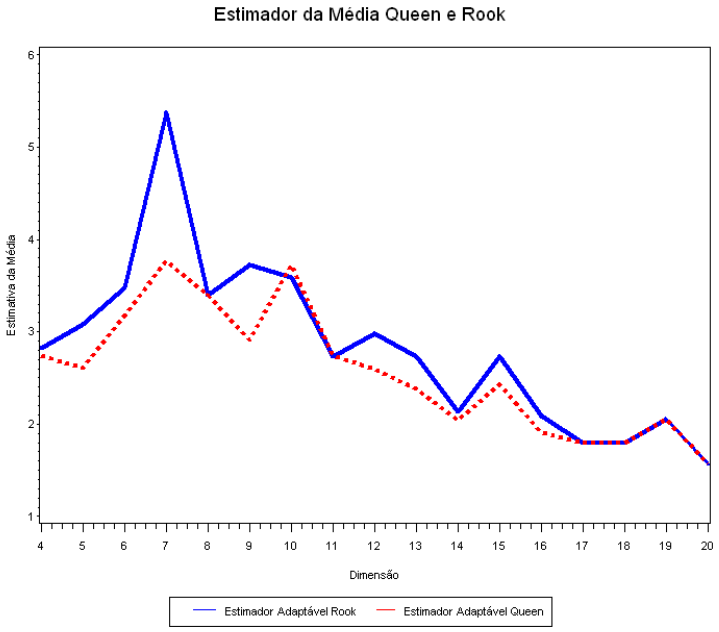


Figura 4.13: Análise da Média com seleção *ROOK* e *QUEEN*

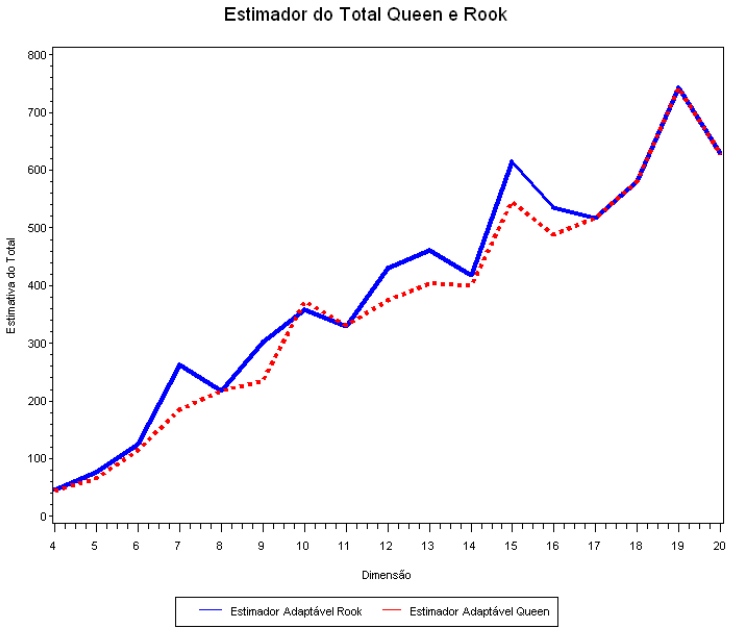


Figura 4.14: Análise do Total com seleção *ROOK* e *QUEEN*

Assim, verifica-se esses estimadores têm valores bem próximos e quando a dimensão é 17 as estimativas da média e do total são as mesmas. Com isso, utilizar a seleção *QUEEN*

ou *ROOK* se torna indiferente.

Na próxima Seção será abordado o exemplo de (Thompson, 2002) para o caso da amostragem espacial adaptável estratificada por conglomerado, bem como os seus estimadores para analisar se há uma diferença significativa ao ser considerado ou não os limites dos estratos.

## 4.5 Amostragem Espacial Adaptável Estratificada por Conglomerado

Thompson (1990) apresenta um exemplo de como funciona a amostragem espacial adaptável estratificada por conglomerados e compara os resultados obtidos quando é considerado os limites entre os estratos ou não. Inicialmente é desenhada uma grade regular em cima da área a ser pesquisada e em seguida são selecionadas $n$ unidades (quadrados) pelo método AAS.

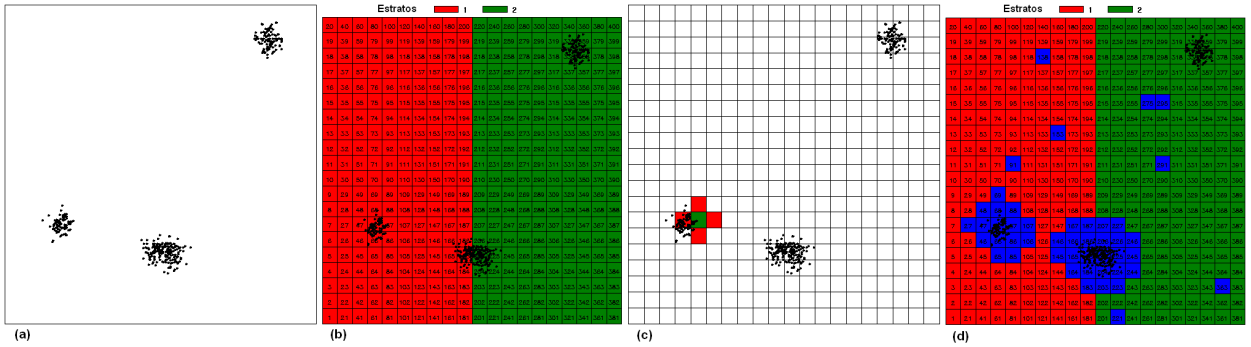


Figura 4.15: Amostragem Espacial Adaptável Estratificada por Conglomerados

O número de objetos encontrados na área analisada na Figura 4.15 (a) é de 397 elementos, dentro de um total de $N = 400$ quadrados. Assim, tem-se que a média populacional é de $\mu = \frac{397}{400} = 0,9925$. Para esse exemplo, a região foi dividida em dois estratos onde uma $AAS$ inicial de $n = 10$ elementos foi selecionada com amostra espacial adaptável estratificada por conglomerado com tamanhos iguais em cada estrato. Dessa forma, no estrato 1 verifica-se um total de $N = 200$ quadrados e $n = 5$ elementos amostrais iniciais; já no estrato 2, encontra-se os outros $N = 200$ e $n = 5$, totalizando a área.

Como no exemplo da amostragem espacial adaptável por conglomerados, a unidade satisfaz a condição se em cada quadrado selecionado for encontrado uma ou mais elementos

de interesse. Sendo verificado essa condição, seleciona-se seus vizinhos. A vizinhança de cada unidade inclui todas as unidades adjacentes. Assim, uma vizinhança pode ser analisada de duas formas: desconsiderando o limite existente entre os estratos ao selecionar os vizinhos de uma unidade; ou considerando esse limite. A seguir será analisado os dois casos.

No primeiro caso, ao ser selecionado uma unidade com elemento e que se encontra no quadrado em contato com a divisão do estrato, esse terá quatro vizinhos: superior, inferior, direita e esquerda, independente de esse se encontrar em um estrato diferente. Com isso, o valor de $w_{hi}$" para o estimador $\widetilde{\mu}$", que ignora o limite dos estratos, é zero para todas as unidades que não satisfazem a condição.

Na primeira rede de intersecção no estrato 1, dada pela Figura 4.15 (d), o valor de $w_{11}" = \frac{96}{6} = 16$. Para a segunda rede de intersecção, o valor é dado por $w_{12}" = \frac{78}{5} = 15,6$, baseado somente nas unidades do estrato 1. Assim, não há intersecções no estrato 2. Portanto, a estimativa da média da população e a estimativa da variância de $\widetilde{\mu}$", dado pelas Equações 2.32 e 2.34, respectivamente, é:

$$\widetilde{\mu}" = \frac{1}{400}\left[\frac{200}{5}(16 + 15,6 + 0 + 0 + 0) + \frac{200}{5}(0 + 0 + 0 + 0 + 0)\right] = 3,16$$

$$\widehat{var}(\widetilde{\mu}") = \frac{1}{400^2}\left[\frac{200(200-5)(74,9)}{5} + 0\right] = 3,65$$

onde $74,9$ é a variância dos cinco números $(16; 15,6; 0; 0; 0)$.

No segundo caso, ao ser selecionado uma unidade com elemento e que se encontra no quadrado em contato com a divisão do estrato, esse terá três vizinhos: superior, inferior, direita (ou esquerda), dependendo de esse se encontrar em um estrato diferente. Com isso, para calcular o estimador $\widetilde{\mu}$ (2.29), usa-se para o mesmo estrato $\frac{n_h}{N_h}$.

Dessa forma, obtem-se as variáveis $w_{hi}$, ou seja, $w_{11} = \frac{96}{6} = 16$ para a primeira rede e $w_{12} = \frac{192}{11} = 17,45$ para a segunda. Assim, a estimativa para a média e sua variância, dada pelas Equações 2.29 e 2.30, respectivamente, é:

$$\widetilde{\mu} = \frac{1}{400}\left[\frac{200}{5}(16 + 17,45 + 0 + 0 + 0) + 0\right] = 3,35$$

$$\widehat{var}(\widetilde{\mu}) = \frac{1}{400^2}\left[\frac{200(200-5)(84,2)}{5} + 0\right] = 4,10$$

onde $84,2$ é a variância dada pelos cinco valores de $w_{1i}$.

A Tabela 4.8 apresenta as estimativas encontradas e a Figura 4.16 mostra a saída do

SAS e verifica-se que não há uma diferença significativa entre os estimadores da média, bem como das variâncias estimadas da média se for considerado os limites dos estratos.

Tabela 4.8: Tabela de Comparação dos Estimadores da Amostragem Espacial Estratificada Adaptável por Conglomerado

| **Estimadores** | ***No Crossing Stratum Boundaries*** | ***Crossing Stratum Boundaries*** |
|---|---|---|
| **Estimadores da Média** | 3,16 | 3,35 |
| **Variância Estimada da Média** | 3,65 | 4,10 |

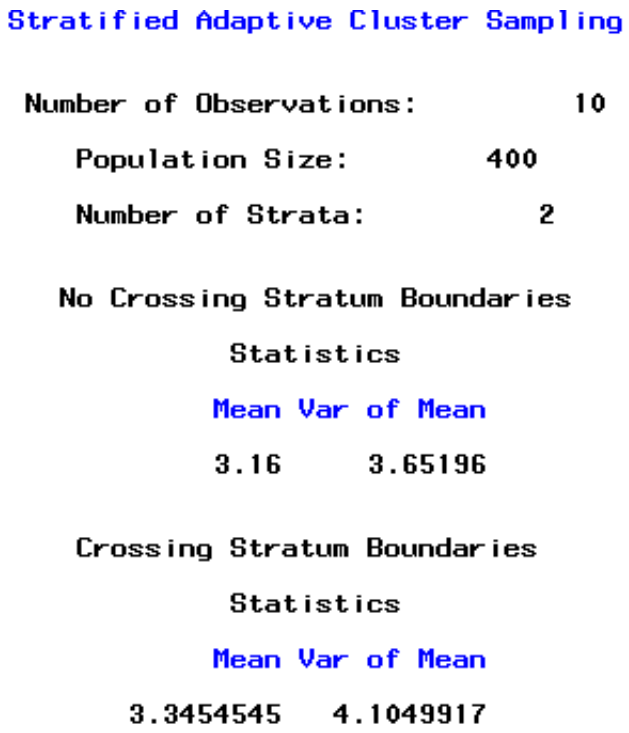


Figura 4.16: Saída do Exemplo da Amostragem Adaptável Estratificada por Conglomerado

# Capítulo 5

# Conclusões

## 5.1 Conclusões

A partir da teoria de amostragem dada no Capítulo 1, foi possível ter a base para entender a teoria da amostragem espacial adaptável do Capítulo 2. Já no Capítulo 3 foi desenvolvido o algoritmo computacional no *software* SAS para a amostragem espacial adaptável por conglomerado e para a amostragem espacial adaptável estratificada por conglomerado, descrita no Capítulo 2. Para isso, algumas simulações foram feitas no *software* SAS com base nos dados dos exemplos do (Thompson, 2002), o que foi abordado no Capítulo 4.

Dessa forma, verificou-se que a amostragem espacial adaptável por conglomerado é a que sofreu menor variação entre as $AAS$ e $AAS$ adaptável, em que foi demonstrado que é uma estimativa viesada. Na Seção 4.3 foi feita a comparação entre os diferentes tamanhos da população, que indicou que a amostragem espacial adaptável por conglomerado sofre menos variações em seus estimadores do que as outras duas amostragens.

Além disso, o estimador $\widetilde{\mu}$ é semelhante tanto para a vizinhança com seleção $ROOK$ quanto para a $QUEEN$. Assim, aquele estimador é robusto para a amostragem espacial adaptável por conglomerado ao ser ponderado por $N$. Similarmente, na Seção 4.4 foi feita a comparação entre os diferentes tamanhos da população com seleção $QUEEN$, obtendo o mesmo resultado que a seleção $ROOK$. A amostragem espacial adaptável foi a que menos variou entre as outras duas para a análise de dados raros como a localização de minérios, por exemplo.

A amostragem espacial adaptável estratificada por conglomerado mostrou que não

há uma diferença significativa entre os estimadores da média, bem como das variâncias estimadas da média se for considerado os limites dos estratos ou desconsiderá-los. Por fim, uma outra contribuição foi feita ao ser analisada a forma de vizinhança com seleção *QUEEN*, verificando que o estimador $\widetilde{\mu}$ é robusto para a amostragem espacial adaptável por conglomerado.

Conclui-se, então, que a implementação computacional da amostragem espacial adaptável realizada neste trabalho é importante, uma vez que essa nova técnica pode ter diversas aplicações e os usuários ainda não tinham uma ferramenta computacional para utilizá-la.

## 5.2 Limitações do Trabalho

O desenvolvimento dos algoritmos necessários para operacionalizar a amostragem espacial adaptável levou um tempo considerável, e por isso nao foi possível utilizar um estudo de caso com dados reais, a fim de melhor demonstrar o potencial dessa técnica. Além dessa limitação, não foi possível implementar todos os estimadores sugeridos por (Thompson, 2002). No entanto, aqueles implementados sao os mais utilizados na prática.

## 5.3 Recomendações para Trabalhos Futuros

Recomenda-se para trabalhos futuros utilizar estudos de cado reais, a implementação computacional dos outros tipos de amostragem espacial adaptável como: a detectabilidade incompleta (Thompson e Seber, 1994); a amostragem espacial adaptável baseada em estatísticas de ordem (Thompson, 1996); amostragem espacial adptável por conglomerados baseada em estatística de ordem (Wald, 1947; Robbins, 1952; Zacks, 1970; Siegmund, 1985; Francis, 1991; Cochran, 1977; Thompson, 2002); a análise multivariada da amostragem espacial adaptável (Bethel; Kokan e Khan, 1967; Cochran, 1977; Thompson, 1993), bem como o desenvolvimento de intervalo de confiança e uma análise do viés do estimador $\overline{y}_{AD}$.

# Referências Bibliográficas


Barry, D. A. & Fristedt, B. Bandit problems: Sequential allocation of experiments. *Biometrical Journal*, 29(1):20.

Basu, D. (1961-2002). Role of the sufficiency and likelihood principles in simple survey theory. *Sankhya: The Indian Journal of Statistics*, 31(4):441 – 454.

Bethel, J. On sample allocation in multivariate surveys. *Survey Methodology*, 15:47 – 57.

Bolfarine, H. & Bussab, W. O. (2005). *Elementos de Amostragem*, (1st ed.). Blucher.

Brown, J. A. *The relative efficiency of adaptive cluster sampling.* PhD thesis.

Brown, J. A. (1994). The application of adaptive cluster sampling to ecological studies. *Statistics in ecology and environmental monitoring*, (2):86 – 97.

Brown, J. A. (1996). The relative efficiency of adaptive cluster sampling for ecological surveys. *Faculty of Inforniation and Mathematical Sciences*, (2):1371–7637.

Brown, J. A., M., M. S., Moradi, M., Bell, G., & Smith, D. R. (2008). An adaptive two-stage sequential design for sampling rare and clusted populations. *The Society of Population Ecology and Springer*, 50:239 – 245.

Chang, M. (2008). *Adaptive Design Theory and Implementation Using SAS and R.* Chapman and Hall/CRC Biostatistics Series.

Cochran, W. G. (1977). *Sampling Techniques*, (3rd ed.). Wiley.

Cornfield, J. (1944). On samples from finite populations. *Journal of the American Statistical Association*, 39(226):236 – 239.

Domingo, C., Gavald , R., & Watanabe, O. (2002). Adaptive sampling methods for scaling up knowledge discovery algorithms. *Discovery Science Lecture Notes in Computer Science*, 6(2):131 – 152.

Feller, W. (1957). *An Introduction to probability theory and its applications*, (2sd ed.). John Wiley.

Foreman, E. K. (1991). *Survey Sampling Principles*, (3rd ed.). Marcel Dekker.

Francis, R. I. C. C. (1991). Statistical properties of two-phase surveys: comment. *Canadian Journal of Fisheries and Aquatic Sciences*, 48:1228.

Hansen, M. H. & Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *Institute of Mathematical Statistics*, 14(4):333 – 362.

Jain, A. & Chang, E. Y. (2004). Adaptive sampling for sensor networks. *Proceedings of the first workshop on data management for sensor networks*, pages 10 – 16.

Kalton, G. & Anderson, D. (1986). Sampling rare populations. *Journal of the Royal Statistical Society*, 149(1):65 – 82.

Khan, A. & Muttlak, H. A. (2002). Adjusted two-stage adaptive cluster sampling. *Environmental and Ecological Statistics*, 9:111 – 120.

Kish, L. (1965). *Survey Sampling*. Wiley.

Kokan, A. R. & Khan, S. (1967). Optimun allocation in multivariate surveys: An analytical solution. *Journal of the Royal Statistical Society*, 29(1):115 – 125.

Kunigami, G. (2010). Ponto dentro de polígono. Technical report, Unicamp.

Lo, N. C. H., Griffith, D., & Hunter, J. R. (1997). Using a restricted adaptive cluster sampling to estimate pacific hake larval abundance. *California Cooperative Oceanic Fisheries Investigations*, 38:103–113.

Lohr, S. L. (1999). *Sampling: Design and Analysis*. Duxbury Press.

Mier, K. L. & Picquelle, S. J. (2008). Estimating abundance of spatially aggregated populations: comparing adaptive sampling with other survey desing. *Canadian Journal of Fisheries & Aquatic Sciences*, 65(2):176–197.

Raj, D. (1968). *Sampling Theory*. McGraw-Hill.

Ramsey, S. K. T. F. L. & Seber, G. A. F. (1992). An adaptive procedure for sampling animal populations. *International Biometric Society*, 48(4):1195 – 1199.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58:527 – 535.

Robinson, A. (2009). Adaptive design theory and implementation using sas and r. *Journal of Applied Statistics*, 36(6):701–702.

Salehi, M. M. & Seber, G. A. F. (1997). Two-stage adaptive cluster sampling. *International Biometrics Society*, 53(3):959 – 970.

Satyanarayana, A. & Davidson, I. A dynamic adaptive sampling algorithm (dasa) for real world applications: Finger print recognition and face recognition. *Foundations of Intelligent Systems*, (3488):631 – 640.

Scheaffer, R. L., III, W. M., & Ott, R. L. (1996). *Elementary Survey Sampling*, (5th ed.). Duxbury.

Seber, G. A. F. (1986). A review of estimating animal abundance. *International Biometric Society*, 42(2):267 – 292.

Seber, G. A. F. & Salehi, M. M. (2013). *Adaptive Sampling Designs: Inference for Sparce and Clustered Populations*, (1th ed.). Springer.

Sengupta, R. N. & Sengupta, A. (2011). Some variants of adaptive sampling procedures and their applications. *Computational Statistics and Data Analysis*, 55:3183 – 3196.

Siegmund, D. (1985). *Sequential Analysis; Tests and Confidence Intervals*. Springer Series in Statistics.

Stein, A. & Ettema, C. (2003). An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. *Agriculture, Ecosystems and Environment*, 94(1):31 – 47.

Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association*, 85(412):1050 – 1059.

Thompson, S. K. (1991). Stratified adaptive cluster sampling. *Biometrika Trust*, 78(2):389–397.

Thompson, S. K. (1993). Multivariate aspects of adaptive cluster sampling. *North-Holland Series in Statistics and Probability/Elsevier Science Publiser*, 6:561 – 572.

Thompson, S. K. (1996). Adaptive cluster sampling based on order statistics. *Environmetrics*, 7(2):123–133.

Thompson, S. K. (2002). *Sampling*, (2sd ed.). Wiley.

Thompson, S. K. (2011). Adaptive sampling. Technical report.

Thompson, S. K. & Horvitz, D. G. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

Thompson, S. K. & Seber, G. A. F. (1994). Detectability in conventional and adaptive sampling. *International Biometric Society*, 50(3):712 – 724.

Thompson, S. K. & Seber, G. A. F. (1996). *Adaptive sampling*. Wiley.

Tukey, J. W. (1950). Some sampling simplified. *Journal of American Statistics Association*, 45(252):501 – 519.

Vieira, M. T. F. A. S. (2008). Amostragem. Master's thesis, Universidade de Aveiro.

Wald, A. (1947). *Sequential Analysis*. Wiley.

Waldispühl, J. & Ponty, Y. (2011). An unbiased adaptive sampling algorithm for the exploration of rna mutational landscapes under evolutionary pressure. *Research in Computational Molecular Biology Lecture Notes in Computer Science*, 6577:501–515.

Woodby, D. (1998). Adaptive cluster sampling: efficiency, fixed sample sizes, and an application to red sea urchins (*Strongylocentrotus franciscanus*) in southeast alaska. *Canadian Specia Publication of Fisheries and Aquatatic Sciences*, (125):15 – 20.

Zacks, S. (1970). Bayesian design for single and double stratified sampling for estimating proportion in finite population. *Technometrics*, 12(1):119 – 130.

# Apêndice A

# Módulos - SAS

```
/************** creating grid ******************/
%macro grid(minx=,maxx=,miny=,maxy=,dim=,anno=,printN=YES);
 %let byx=%sysfunc(int(%sysevalf(100*(&maxx-&minx)/&dim)));%put &byx;
 %let byy=%sysfunc(int(%sysevalf(100*(&maxy-&miny)/&dim)));%put &byy;
 %let minx=%sysevalf(100*&minx);
 %let maxx=%sysevalf(100*&maxx);
 %let miny=%sysevalf(100*&miny);
 %let maxy=%sysevalf(100*&maxy);
  proc sql;
   create table trab&dim (x num, y num);
   %do i=&minx %to &maxx %by &byx;
    %do j=&miny %to &maxy %by &byy;
     %let l=%eval(&i+&byx);
     %let m=%eval(&j+&byy);
     %if &l<=&maxx and &m<=&maxy and &i<=&maxx and &j<=&maxy %then %do;
      insert into trab&dim
      values (&i,&j)
      values (&i,&m)
      values (&l,&m)
      values (&l,&j);
     %end;
    %end;
   %end;
  quit;
data id&dim;
 do id=1 to &dim*&dim;
  do i=1 to 4;
   output;
  end;
 end;
run;
data trab&dim;
 merge trab&dim id&dim;
  x=x/100;
  y=y/100;
 drop i;
run;
proc sql;
 create table coor as
 select distinct id, max(x) as maxx,max(y) as maxy, min(x) as minx,
   min(y) as miny
 from trab&dim
 group by id
 order by id;
quit;
data coor;
 set coor;
  x=(maxx+minx)/2;
  y=(maxy+miny)/2;
  position='5';
  function='label';
  style='simplex';
  text=trim(left(put(id,$4.)));
  size=0.8;
  color='black';
  xsys='2';
  ysys='2';
  when='a';
run;
 %let r=&dim;
 %let n=%eval(&r*&r);
data a&dim;
  %do i=1 %to &n;
   %if &i<=&r or &i>=%eval(&r*(&r-1)+1) or %qsysfunc(mod(&i,&r))=0 or
        %qsysfunc(mod(&i,&r))=1 %then %do;
    id=&i;
    v=0;
    output;
   %end;
   %do l=1 %to (&r-1)/2;
    %do j=1 %to &r;
     %if &i=%eval((&l+1)+(&j*&r)) or &i=%eval((&r-&l)+(&j*&r))or
         &i=%eval((&l*&r)+(&j+&l)) or &i=%eval((((&r-1)-&l)*&r)+(&j+&l)) %then %do;
      id=&i;
      v=&l*10;
      output;
     %end;
     %end;
   %end;
  %end;
run;
%if %upcase(&printN)=YES %then %do;
data &anno._;
 set &anno coor;
run;
%end;
%else %do;
data &anno._;
 set &anno;
run;
%end;
proc sort data=a&dim nodupkey;
 by id;
run;
goptions reset=all reset=global ftitle='Verdana';
data a;id=9999;v=2;run;
title "Grid of Dimension &dim";
proc gmap data=a map=trab&dim all;
 id id;
 choro v / nolegend anno=&anno._;
 run;
quit;
%mend grid;


/************** defining neighborhood ******************/
%macro neighborhood(id=,pt=,map=,anno=,out=,type=ROOK);
proc iml;
use &map;
read all;
ptid=&pt;
read all var{x y} into point where(&id=ptid);
n=nrow(x);
do i=1 to 4;
 do j=1 to n;
```

```
  if point[i,1]=x[j] & point[i,2]=y[j] & ptid ^= &id[j] then do;
    neighbor=neighbor//&id[j];
  end;
 end;
end;
%if %upcase(&type)=ROOK or %upcase(&type)= %then %do;
 call sort(neighbor,{1});
 do i=2 to nrow(neighbor);
  if neighbor[i]=neighbor[i-1] then neighbor2=neighbor2//neighbor[i];
 end;
%end;
%else %do;
 neighbor2=unique(neighbor)';
%end;
neighbor2=neighbor2||j(nrow(neighbor2),1,ptid);
create &out from neighbor2[colname={"&id" "v"}];
append from neighbor2;
quit;
proc sql;
 insert into &out values(&pt,9999);
quit;
goptions reset=all reset=global ftitle='Verdana';
title "Adjacent Neighboring Units of &id=&pt";
proc gmap data=&out map=&map all;
 id &id;
 choro v / nolegend %if &anno ne %then %do; anno=&anno;%end;;
 run;
quit;
%mend neighborhood;



/***** point inside grid *******/
%macro ginside(map=,id=,where=,data=,out=);
proc iml;
use &map;
read all var{x} into _p1;
read all var{y} into _p2;
read all var{&id} into _id;
p=_p1||_p2||_id;
free _p1 _p2 _id;
start verifyisright(p1,p2,point);
isleft=0;
interceptasegmento=0;
numberofinterceptededges=0;
ymin=min(p1[2],p2[2]);
ymax=max(p1[2],p2[2]);
y=point[2];
if (p2[2]-p1[2])^= 0 then x=p1[1]+
  (y-p1[2])*(p2[1]-p1[1])/(p2[2]-p1[2]);
else x=p1[1];
if point[1]<x then isleft=1;
if y>ymin & y<=ymax then interceptasegmento=1;
if isleft=1 & interceptasegmento=1 then numberofinterceptededges=1;
return(numberofinterceptededges);
finish verifyisright;
start edge(p,point);
p1=p[1,];
p2=p[2,];
point=point;
cond=verifyisright(p1,p2,point);
return(cond);
finish edge;
start isinsidepolygon(point) global(p);
%if &where = %then %do;
 id=unique(p[,3]);
%end;
%else %do;
 id=&where;
%end;
do j=1 to ncol(id);
 _type_=0;
 point=point;
 numberofinterceptededges=0;
 pp=p[loc(p[,3]=id[j]),];
 do i=1 to nrow(pp)-1;
  z=edge(pp[i:i+1,],point);
   if z=1 then numberofinterceptededges=numberofinterceptededges+1;
   *print numberofinterceptededges;
 end;
 if mod(numberofinterceptededges,2)=1 then _type_=1;
 if _type_=1 then do;
   result=point||_type_||id[j];
   append from result;
 end;
end;
finish isinsidepolygon;
use &data;
read all into point;
result=j(1,4,0);
create &out from result[colname={"x" "y" "_type_" "id"}];
do j=1 to nrow(point);
 run isinsidepolygon(point[j,]);
end;
close &out;
quit;
%mend ginside;



/***************** adaptive sampling ****************/
%macro as(data=,n=,sample=,out=,strata=,seed=,map=,id=,anno=,typen=ROOK,printN=YES);
  %if &sample= %then %do;
    proc surveyselect data=&data sampsize=&n out=&out seed=&seed noprint;
    run;
    data &out;set &out;
     v=2;
    run;
  %end;
  %else %do;
    data &out;set &sample;
 v=2;
    run;
  %end;
title "Selected Sampling";
proc gmap data=&out map=&map all;
  id &id;
  choro v / nolegend anno=&anno;
run;
quit;
data &anno.1;
 set &anno(keep=x y);
run;
proc sql noprint;
 select count(*) into:n from &out;
quit;
%put &n;
data _null_;
 set &out;
 call symput('id'||trim(left(_n_)),&id);
run;
%put &id1;
proc sql;
 drop table &out.s1,&out.s2,&out.s3,_freq_as_,&out.neighbor2;
 create table _freq_as_ (id num, freq num);
quit;
%do i=1 %to &n;
%ginside(map=&map,id=&id,where=&&id&i,data=&anno.1,out=&out.out);
  %if &i=1 %then %do;
    %let cond=0;
  %end;
  %else %do;
    proc sql noprint;
    select count(*) into:cond from &out.s1 where id in
        (select distinct id from &out.out);
    quit;
  %end;
%put "cond=" &cond;
  %if &cond=0 %then %do;
    proc append base=&out.s1 data=&out.out force;
    run;
  %end;
%end;
proc sql;
```

```
  insert into _freq_as_ select id,count(*) as freq from &out.s1
     group by id;
  quit;
proc sort data=&out.s1 nodupkey;
 by &id;
run;
proc sql noprint;
  select count(*) into:m from &out.s1;
quit;
%put &m;
data _null_;set &out.s1;
 call symput('ids'||trim(left(_n_)),&id);
 run;
 %put &ids1;
%do j=1 %to &m;
 %neighborhood(id=&id,pt=&&ids&j,map=&map,anno=&anno,
    out=&out.neighbor,type=&typen);
  proc append base=&out.s2 data=&out.neighbor force;
  run;
  proc append base=&out.neighbor2 data=&out.neighbor force;
  run;
%end;
data &out.s2;
  set &out.s2;
  v=2;
run;
proc append base=&out.s3 data=&out.s2 force;
run;
proc sql noprint;
 create table &out.s2 as select * from &out.s2
 where &id not in (select &id from &out);
 select count(*) into:k from &out.s2;
quit;
%put &k;
data _null_;
 set &out.s2;
 call symput('id'||trim(left(_n_)),&id);
 run;
%put &id1;
%do %while (&k>0);
 proc sql;
   drop table &out.s1,&out.s2;
 quit;
%do i=1 %to &k;
 %ginside(map=&map,id=&id,where=&&id&i,data=&anno.1,out=&out.out);
 %if &i=1 %then %do;
   %let cond=0;
 %end;
 %else %do;
   proc sql noprint;
    select count(*) into:cond from &out.s1 where id in
        (select distinct id from &out.out);
   quit;
 %end;
%put "cond=" &cond;
 %if &cond=0 %then %do;
   proc append base=&out.s1 data=&out.out force;
   run;
 %end;
%end;
proc sql;
  insert into _freq_as_ select id,count(*) as freq from &out.s1
     group by id;
quit;
proc sort data=&out.s1 nodupkey;
 by &id;
run;
proc sql noprint;
  select count(*) into:m from &out.s1;
quit;
%put &m;
%if &m>0 %then %do;
 data _null_;
  set &out.s1;
  call symput('ids'||trim(left(_n_)),&id);
  run;
%put &ids1;
 %do j=1 %to &m;
  %neighborhood(id=&id,pt=&&ids&j,map=&map,anno=&anno,
     out=&out.neighbor,type=&typen);
  proc append base=&out.s2 data=&out.neighbor force;
  run;
  proc append base=&out.neighbor2 data=&out.neighbor force;
  run;
 %end;
%end;
data &out.s2;set &out.s2;
 v=2;
run;
proc append base=&out.s3 data=&out.s2 force;
run;
proc sql noprint;
%if &strata= %then %do;
 insert into &out select &id,2 from &out.s1;
%end;
%else %do;
insert into &out select &id,2,0 from &out.s1;
%end;
 create table &out.s2 as select * from &out.s2
 where &id not in (select &id from &out);
 select count(*) into:k from &out.s2;
quit;
%put k=&k;
 data _null_;set &out.s2;
   call symput('id'||trim(left(_n_)),&id);
 run;
 %put &id1;
%end;
proc sort data=&out.s3 nodupkey;
 by &id;
run;
proc sort data=_freq_as_;
 by id;
run;
data &out.s3;
 set &out.s3 &out;
run;
proc sort data=&out.s3 nodupkey;
 by &id;
run;
%if %upcase(&printN)=YES %then %do;
data &anno._;
 set &anno coor;
run;
%end;
%else %do;
data &anno._;
 set &anno;
run;
%end;
%if &strata= %then %do;
goptions reset=all reset=global ftitle='Verdana';
title "Adaptive Cluster Sampling";
proc gmap data=&out.s3 map=&map all;
 id &id;
 choro v / nolegend anno=&anno._;
run;
quit;
%end;
%else %do;
data &out.s3;set &out.s3(in=a) &data;
 if a then strata=99999;
run;
proc sql noprint;select max(strata) into:mstrata from &data;quit;%put &mstrata;
title "Stratified Adaptive Cluster Sampling";
proc gmap data=&out.s3 map=&map all;
 id &id;
 choro strata / anno=&anno._ legend=legend1;
 legend1 order=(1 to &mstrata);
run;
quit;
%end;
```

```
proc freq data=&out.neighbor2 noprint;
 tables v /out=_sample_(drop=count percent);
run;
data &out.neighbor2;set &out.neighbor2;
 source=0;
 if v ne 9999;
run;
data &out;set &out;
 seq=_n_;
run;
proc sql noprint;
 create table _sample_ as
 select v as id from _sample_ where v in (select id from &out
   where seq<=&n);
 update &out.neighbor2 set source=v where v in (select id from
  _sample_);
 create table _refresh_ as
 select id as v,source from &out.neighbor2
 where v in (select id from &out.neighbor2) and source ne 0;
 create table &out.neighbor3 as
 select a.*,b.source as s2 from &out.neighbor2 a left join
  _refresh_ b on a.v=b.v;
 update &out.neighbor3 set source=s2 where s2 ne .;
 alter table &out.neighbor3 drop s2;
 select count(*) into:zero from &out.neighbor3 where source=0;
 %put &zero;
 %do %while (&zero>0);
  create table _refresh_ as
  select distinct id as v,source from &out.neighbor3
  where v in (select id from &out.neighbor3) and source ne 0;
  create table &out.neighbor3 as
  select a.*,b.source as s2 from &out.neighbor3 a left join
   _refresh_ b on a.v=b.v;
  update &out.neighbor3 set source=s2 where s2 ne .;
  alter table &out.neighbor3 drop s2;
  select count(*) into:zero from &out.neighbor3 where source=0;
  %put &zero;
 %end;
quit;
proc sort data=&out.neighbor3(drop=v) nodupkey;
 by id source;
run;
data _freq_as_;
 merge _freq_as_ &out.neighbor3;
 by id;
run;
proc sort data=&out;
 by id;
run;
data &out;retain id freq;
 merge &out(in=a) _freq_as_;
 by id;
 if freq=. then freq=0;
 if source=. then source=0;
run;
proc sort data=&out nodupkey;
 by id;
run;
data &out;retain &id freq v seq;
 merge &out(in=a) &data(keep=id &strata);
 by id;
 if a;
run;
%if &strata= %then %do;
proc sort data=&out;
 by source;
run;
%end;
%else %do;
proc sort data=&out;
 by &strata source;
run;
%end;
proc iml;
 use &out;
 read all into tab;
 use &data;
 read all into pop;
 NN=nrow(pop);
 n=&n;
tab1=tab[loc(tab[,4]^=.),];
*print tab1;
%if &strata= %then %do;
 mu=j(1,3,1);
 mu[1]=tab1[1,ncol(tab1)];
 mu[2]=tab1[1,2];
  do j=2 to nrow(tab1);
   if tab1[j,ncol(tab1)]=tab1[j-1,ncol(tab1)] then do;
      mu[nrow(mu),2]=mu[nrow(mu),2]+tab1[j,2];
      mu[nrow(mu),ncol(mu)]=mu[nrow(mu),ncol(mu)]+1;
   end;
   else do;
      mu=mu//(tab1[j,ncol(tab1)]||tab1[j,2]||j(1,1,1));
   end;
  end;
  *print mu;
 mu=mu||choose(mu[,2]=0, mu[,3], 1);
 u1=(1/n)*(mu[,2]/mu[,3])`*j(nrow(mu),1,1);
 varu1=(NN-n)/(NN*n*(n-1))*((mu[,4]#((mu[,2]/mu[,3])-u1))`*((mu[,2]/mu[,3])-u1));
 Totu1=NN*u1;
 TotVaru1=NN**2*varu1;
 print "Number of Observations: " n[label=' '],, "Population Size: " NN[label=' '];
 Print "Adaptive Cluster Sampling",,"Statistics",, u1[label='Mean']
    varu1[label='Var of Mean'] Totu1[label='Sum'] TotVaru1[label='Var of Sum'];
*SRS;
fsample=tab[loc(tab[,4]<=n),1:2];
SRS=fsample[+,2]/n;
VARSRS=((1-n/NN)/n)*((fsample[,2]-SRS)`*(fsample[,2]-SRS)/(n-1));
TotSRS=NN*SRS;
TotVarSRS=NN**2*varSRS;
Print "Simple Random Sampling",,"Statistics",, SRS[label='Mean']
    VARSRS[label='Var of Mean'] TotSRS[label='Sum'] TotVarSRS[label='Var of Sum'];
*adapSRS;
adSRS=tab[+,2]/nrow(tab);
VARadSRS=((1-nrow(tab)/NN)/nrow(tab))*((tab[,2]-adSRS)`*(tab[,2]-adSRS)/(nrow(tab)-1));
TotadSRS=NN*adSRS;
TotVaradSRS=NN**2*VARadSRS;
Print "Adaptive Simple Random Sampling (biased)",,"Statistics",, adSRS[label='Mean']
    VARadSRS[label='Var of Mean'] TotadSRS[label='Sum'] TotVaradSRS[label='Var of Sum'];
create _estimates_&map._ var{u1 varu1 Totu1 TotVaru1 SRS VARSRS TotSRS TotVarSRS
    adSRS VARadSRS TotadSRS TotVaradSRS};
append;
%end;
%else %do;


/*** no crossing stratum boundaries ****/
  NNh=j(1,2,1);NNh[1]=pop[1,ncol(pop)];
  do j=2 to nrow(pop);
   if pop[j,ncol(pop)]=pop[j-1,ncol(pop)] then NNh[nrow(NNh),2]=NNh[nrow(NNh),2]+1;
   else NNh=NNh//(pop[j,ncol(pop)]||j(1,1,1));
  end;
  *print NNh;
  nstrata=NNh[<>,1];
  print "Number of Observations: " n[label=' '],, "Population Size: " NN[label=' '],,
    "Number of Strata: " nstrata[label=' '];
  fsample=tab1[loc(tab1[,4]<=n),5];
  nh=j(1,2,1);
  nh[1]=fsample[1];
  do j=2 to nrow(fsample);
   if fsample[j]=fsample[j-1] then nh[nrow(nh),2]=nh[nrow(nh),2]+1;
   else nh=nh//(fsample[j]||j(1,1,1));
  end;
  *print nh;
  mu=j(1,4,1);
  mu[1]=tab1[1,ncol(tab1)-1];
  mu[2]=tab1[1,ncol(tab1)];
  mu[3]=tab1[1,2];
  do j=2 to nrow(tab1);
   if tab1[j,ncol(tab1)-1]=tab1[j-1,ncol(tab1)-1] &
    tab1[j,ncol(tab1)]=tab1[j-1,ncol(tab1)] then do;
     mu[nrow(mu),3]=mu[nrow(mu),3]+tab1[j,2];
```

```
    mu[nrow(mu),ncol(mu)]=mu[nrow(mu),ncol(mu)]+1;
  end;
  else do;
    mu=mu//(tab1[j,ncol(tab1)-1]||tab1[j,ncol(tab1)]||
     tab1[j,2]||j(1,1,1));
  end;
 end;
 *print mu;
 mu=mu||mu[,3]/mu[,4]||choose(mu[,3]=0,
   mu[,4], 1)||nh[mu[,1],2]||NNh[mu[,1],2]||j(nrow(mu),1,0);
 do k=1 to nrow(mu);
  if mu[k,2]^=0 then do;
   if tab1[loc(tab1[,1]=mu[k,2]),5]^=mu[k,1] then mu[k,5]=0;
  end;
 end;
 kk=1;
 var=0;
 do j=2 to nrow(mu);
  if mu[j,1]=mu[j-1,1] then var=var+mu[j,5];
  else do;
    mu[kk:j-1,ncol(mu)]=repeat(var/mu[j-1,ncol(mu)-2],j-kk);
    kk=j;
    var=mu[j,5];
  end;
  if j=nrow(mu) then mu[kk:j,ncol(mu)]=repeat(var/mu[j,ncol(mu)-2],j-kk);
 end;
u1=(1/NN)*((mu[,ncol(mu)-1]/mu[,ncol(mu)-2])#mu[,5])'*j(nrow(mu),1,1);
s2h=(1/(mu[,ncol(mu)-2]-1))#mu[,6]#(mu[,5]-mu[,ncol(mu)])##2;
varu1=(1/NN**2)*((mu[,ncol(mu)-1]/mu[,ncol(mu)-2])#
   (mu[,ncol(mu)-1]-mu[,ncol(mu)-2])#s2h)'*j(nrow(mu),1,1);
print "No Crossing Stratum Boundaries",,"Statistics",, u1[label='Mean']
   varu1[label='Var of Mean'];


/*** crossing stratum boundaries ****/
call sort(tab1,{6,5});
*print tab1;
tab2=tab1[loc(tab1[,6]=0),];
*print tab2;
mu=j(1,3,1);
mu[1]=tab1[1,ncol(tab1)];
mu[2]=tab1[1,2];
do j=2 to nrow(tab1);
 if tab1[j,ncol(tab1)]=tab1[j-1,ncol(tab1)] then do;
   mu[nrow(mu),2]=mu[nrow(mu),2]+tab1[j,2];
   mu[nrow(mu),ncol(mu)]=mu[nrow(mu),ncol(mu)]+1;
 end;
 else do;
   mu=mu//(tab1[j,ncol(tab1)]||tab1[j,2]||j(1,1,1));
 end;
end;
*print mu;
mu=mu[loc(mu[,1]=0)+1:nrow(mu),];
mu2=j(1,4,1);
mu2[1]=tab1[1,ncol(tab1)-1];
mu2[2]=tab1[1,ncol(tab1)];
mu2[3]=tab1[1,2];
do j=2 to nrow(tab2);
 if tab2[j,ncol(tab2)-1]=tab2[j-1,ncol(tab2)-1] &
       tab2[j,ncol(tab2)]=tab2[j-1,ncol(tab2)] then do;
   mu2[nrow(mu2),3]=mu2[nrow(mu2),3]+tab1[j,2];
   mu2[nrow(mu2),ncol(mu2)]=mu2[nrow(mu2),ncol(mu2)]+1;
 end;
 else do;
   mu2=mu2//(tab2[j,ncol(tab2)-1]||tab2[j,ncol(tab2)]||tab2[j,2]||j(1,1,1));
 end;
end;
*print mu2;
mu=j(nrow(mu),1,1)||mu||mu[,2]/mu[,3]||choose(mu[,2]=0, mu[,3], 1);
mu2=mu2||j(nrow(mu2),1,0)||choose(mu2[,3]=0, mu2[,4], 1);
do k=1 to nrow(mu);
 if mu[k,2]^=0 then mu[k,1]=tab1[loc(tab1[,1]=mu[k,2]),5];
end;
mu=mu//mu2;
call sort(mu,{1,2});
mu=mu||nh[mu[,1],2]||NNh[mu[,1],2]||j(nrow(mu),1,0);
```

```
kk=1;
var=0;
do j=2 to nrow(mu);
  if mu[j,1]=mu[j-1,1] then var=var+mu[j,5];
  else do;
    mu[kk:j-1,ncol(mu)]=repeat(var/mu[j-1,ncol(mu)-2],j-kk);
    kk=j;
    var=mu[j,5];
  end;
  if j=nrow(mu) then do;
if kk<j then mu[kk:j,ncol(mu)]=repeat(var/mu[j,ncol(mu)-2],j-kk);
else mu[kk:j,ncol(mu)]=repeat(var/mu[j,ncol(mu)-2],1);
  end;
end;
u1c=(1/NN)*((mu[,ncol(mu)-1]/mu[,ncol(mu)-2])#mu[,5])'*j(nrow(mu),1,1);
s2hc=(1/(mu[,ncol(mu)-2]-1))#mu[,6]#(mu[,5]-mu[,ncol(mu)])##2;
varu1c=(1/NN**2)*((mu[,ncol(mu)-1]/mu[,ncol(mu)-2])#
   (mu[,ncol(mu)-1]-mu[,ncol(mu)-2])#s2hc)'*j(nrow(mu),1,1);
print "Crossing Stratum Boundaries",,"Statistics",,u1c[label='Mean']
   varu1c[label='Var of Mean'];
%end;
quit;
%mend as;
```

# Apêndice B

# Exemplos - SAS

## B.1 Data - Example Adaptive Cluster Sampling

```
*Example Adaptive Cluster Sampling Steven K. Thompson (2002);
data points2;length text $5.;retain hsys '3' xsys ysys '2' when 'a';
function='label';color='black';size=2;style='special';text='J';x=4.1;y=19.2;output;
function='label';color='black';size=2;style='special';text='J';x=4.7;y=19.1;output;
function='label';color='black';size=2;style='special';text='J';x=4.9;y=19.05;output;
function='label';color='black';size=2;style='special';text='J';x=4.85;y=19.2;output;
function='label';color='black';size=2;style='special';text='J';x=4.8;y=19.6;output;
function='label';color='black';size=2;style='special';text='J';x=5.05;y=19.1;output;
function='label';color='black';size=2;style='special';text='J';x=5.06;y=19.15;output;
function='label';color='black';size=2;style='special';text='J';x=5.07;y=19.18;output;
function='label';color='black';size=2;style='special';text='J';x=5.08;y=19.2;output;
function='label';color='black';size=2;style='special';text='J';x=5.09;y=19.22;output;
function='label';color='black';size=2;style='special';text='J';x=5.1;y=19.23;output;
function='label';color='black';size=2;style='special';text='J';x=5.7;y=19.23;output;
function='label';color='black';size=2;style='special';text='J';x=5.75;y=19.22;output;
function='label';color='black';size=2;style='special';text='J';x=5.80;y=19.2;output;
function='label';color='black';size=2;style='special';text='J';x=5.85;y=19.18;output;
function='label';color='black';size=2;style='special';text='J';x=5.90;y=19.15;output;
function='label';color='black';size=2;style='special';text='J';x=5.95;y=19.1;output;
function='label';color='black';size=2;style='special';text='J';x=5.5;y=19.7;output;
function='label';color='black';size=2;style='special';text='J';x=6.05;y=19.1;output;
function='label';color='black';size=2;style='special';text='J';x=6.08;y=19.08;output;
function='label';color='black';size=2;style='special';text='J';x=6.1;y=19.4;output;
function='label';color='black';size=2;style='special';text='J';x=4.9;y=18.7;output;
function='label';color='black';size=2;style='special';text='J';x=4.8;y=18.9;output;
function='label';color='black';size=2;style='special';text='J';x=5.5;y=18.9;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=5.4;y=18.7;output;
function='label';color='black';size=2;style='special';text='J';x=5.5;y=18.45;output;
function='label';color='black';size=2;style='special';text='J';x=5.35;y=18.1;output;
function='label';color='black';size=2;style='special';text='J';x=5.1;y=18.65;output;
function='label';color='black';size=2;style='special';text='J';x=5.2;y=18.78;output;
function='label';color='black';size=2;style='special';text='J';x=5.25;y=18.95;output;
function='label';color='black';size=2;style='special';text='J';x=5.8;y=18.9;output;
function='label';color='black';size=2;style='special';text='J';x=5.85;y=18.92;output;
function='label';color='black';size=2;style='special';text='J';x=5.89;y=18.95;output;
function='label';color='black';size=2;style='special';text='J';x=5.95;y=18.98;output;
function='label';color='black';size=2;style='special';text='J';x=6.05;y=18.8;output;
function='label';color='black';size=2;style='special';text='J';x=6.01;y=18.3;output;
function='label';color='black';size=2;style='special';text='J';x=9.3;y=6.09;output;
function='label';color='black';size=2;style='special';text='J';x=9.02;y=6.1;output;
function='label';color='black';size=2;style='special';text='J';x=9.8;y=6.2;output;
function='label';color='black';size=2;style='special';text='J';x=10.2;y=6.3;output;
function='label';color='black';size=2;style='special';text='J';x=8.48;y=5.66;output;
function='label';color='black';size=2;style='special';text='J';x=8.65;y=5.3;output;
function='label';color='black';size=2;style='special';text='J';x=8.75;y=5.32;output;
function='label';color='black';size=2;style='special';text='J';x=8.91;y=5.15;output;
function='label';color='black';size=2;style='special';text='J';x=8.91;y=5.99;output;
function='label';color='black';size=2;style='special';text='J';x=8.1;y=4.4;output;
function='label';color='black';size=2;style='special';text='J';x=8.2;y=4.6;output;
function='label';color='black';size=2;style='special';text='J';x=8.93;y=4.3;output;
function='label';color='black';size=2;style='special';text='J';x=8.93;y=4.2;output;
function='label';color='black';size=2;style='special';text='J';x=8.93;y=4.2;output;
function='label';color='black';size=2;style='special';text='J';x=7.81;y=3.53;output;
function='label';color='black';size=2;style='special';text='J';x=7.92;y=3.98;output;
function='label';color='black';size=2;style='special';text='J';x=9.1;y=3.98;output;
function='label';color='black';size=2;style='special';text='J';x=9.3;y=3.98;output;
function='label';color='black';size=2;style='special';text='J';x=9.09;y=3.52;output;
function='label';color='black';size=2;style='special';text='J';x=8.07;y=3.52;output;
function='label';color='black';size=2;style='special';text='J';x=8.32;y=3.16;output;
function='label';color='black';size=2;style='special';text='J';x=8.71;y=3.42;output;
function='label';color='black';size=2;style='special';text='J';x=8.71;y=3.52;output;
function='label';color='black';size=2;style='special';text='J';x=8.69;y=3.79;output;
function='label';color='black';size=2;style='special';text='J';x=8.95;y=3.42;output;
function='label';color='black';size=2;style='special';text='J';x=8.89;y=3.31;output;
function='label';color='black';size=2;style='special';text='J';x=8.09;y=3.79;output;
function='label';color='black';size=2;style='special';text='J';x=8.13;y=3.45;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=8.69;y=3.92;output;
function='label';color='black';size=2;style='special';text='J';x=8.21;y=3.76;output;
function='label';color='black';size=2;style='special';text='J';x=8.23;y=3.78;output;
function='label';color='black';size=2;style='special';text='J';x=8.35;y=3.94;output;
function='label';color='black';size=2;style='special';text='J';x=8.22;y=3.82;output;
function='label';color='black';size=2;style='special';text='J';x=8.51;y=3.92;output;
function='label';color='black';size=2;style='special';text='J';x=8.61;y=3.94;output;
function='label';color='black';size=2;style='special';text='J';x=8.71;y=3.93;output;
function='label';color='black';size=2;style='special';text='J';x=8.71;y=3.91;output;
function='label';color='black';size=2;style='special';text='J';x=8.89;y=3.85;output;
function='label';color='black';size=2;style='special';text='J';x=8.94;y=3.93;output;
function='label';color='black';size=2;style='special';text='J';x=8.95;y=3.94;output;
function='label';color='black';size=2;style='special';text='J';x=8.90;y=3.89;output;
function='label';color='black';size=2;style='special';text='J';x=9.25;y=4.05;output;
function='label';color='black';size=2;style='special';text='J';x=9.13;y=4.31;output;
function='label';color='black';size=2;style='special';text='J';x=9.12;y=4.94;output;
function='label';color='black';size=2;style='special';text='J';x=9.21;y=4.15;output;
function='label';color='black';size=2;style='special';text='J';x=9.21;y=4.93;output;
function='label';color='black';size=2;style='special';text='J';x=9.25;y=4.95;output;
function='label';color='black';size=2;style='special';text='J';x=9.55;y=4.89;output;
function='label';color='black';size=2;style='special';text='J';x=9.58;y=4.91;output;
function='label';color='black';size=2;style='special';text='J';x=9.76;y=4.76;output;
function='label';color='black';size=2;style='special';text='J';x=9.76;y=4.89;output;
function='label';color='black';size=2;style='special';text='J';x=9.75;y=4.95;output;
function='label';color='black';size=2;style='special';text='J';x=9.80;y=4.92;output;
function='label';color='black';size=2;style='special';text='J';x=9.95;y=4.95;output;
function='label';color='black';size=2;style='special';text='J';x=10.12;y=4.51;output;
function='label';color='black';size=2;style='special';text='J';x=10.11;y=4.50;output;
function='label';color='black';size=2;style='special';text='J';x=10.08;y=4.96;output;
function='label';color='black';size=2;style='special';text='J';x=10.09;y=4.93;output;
function='label';color='black';size=2;style='special';text='J';x=10.51;y=5.51;output;
function='label';color='black';size=2;style='special';text='J';x=10.71;y=5.68;output;
function='label';color='black';size=2;style='special';text='J';x=10.16;y=5.03;output;
function='label';color='black';size=2;style='special';text='J';x=10.24;y=5.48;output;
function='label';color='black';size=2;style='special';text='J';x=10.17;y=5.36;output;
function='label';color='black';size=2;style='special';text='J';x=10.14;y=5.73;output;
function='label';color='black';size=2;style='special';text='J';x=10.15;y=5.78;output;
function='label';color='black';size=2;style='special';text='J';x=10.16;y=5.83;output;
function='label';color='black';size=2;style='special';text='J';x=10.21;y=5.62;output;
function='label';color='black';size=2;style='special';text='J';x=10.28;y=5.69;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=9.51;y=5.51;output;
function='label';color='black';size=2;style='special';text='J';x=9.52;y=5.52;output;
function='label';color='black';size=2;style='special';text='J';x=9.85;y=5.85;output;
function='label';color='black';size=2;style='special';text='J';x=9.86;y=5.86;output;
function='label';color='black';size=2;style='special';text='J';x=9.12;y=5.83;output;
function='label';color='black';size=2;style='special';text='J';x=9.83;y=5.82;output;
function='label';color='black';size=2;style='special';text='J';x=9.87;y=5.97;output;
function='label';color='black';size=2;style='special';text='J';x=9.88;y=5.98;output;
function='label';color='black';size=2;style='special';text='J';x=9.87;y=5.97;output;
function='label';color='black';size=2;style='special';text='J';x=9.12;y=5.62;output;
function='label';color='black';size=2;style='special';text='J';x=9.11;y=5.61;output;
function='label';color='black';size=2;style='special';text='J';x=9.22;y=5.32;output;
function='label';color='black';size=2;style='special';text='J';x=9.23;y=5.33;output;
function='label';color='black';size=2;style='special';text='J';x=9.25;y=5.53;output;
function='label';color='black';size=2;style='special';text='J';x=9.24;y=5.34;output;
function='label';color='black';size=2;style='special';text='J';x=9.26;y=5.54;output;
function='label';color='black';size=2;style='special';text='J';x=9.37;y=5.45;output;
function='label';color='black';size=2;style='special';text='J';x=9.27;y=5.66;output;
function='label';color='black';size=2;style='special';text='J';x=9.28;y=5.49;output;
function='label';color='black';size=2;style='special';text='J';x=9.32;y=5.75;output;
function='label';color='black';size=2;style='special';text='J';x=9.33;y=5.57;output;
function='label';color='black';size=2;style='special';text='J';x=9.34;y=5.65;output;
function='label';color='black';size=2;style='special';text='J';x=9.46;y=5.71;output;
function='label';color='black';size=2;style='special';text='J';x=9.38;y=5.65;output;
function='label';color='black';size=2;style='special';text='J';x=9.36;y=5.76;output;
function='label';color='black';size=2;style='special';text='J';x=9.38;y=5.67;output;
function='label';color='black';size=2;style='special';text='J';x=9.53;y=5.24;output;
function='label';color='black';size=2;style='special';text='J';x=9.63;y=5.23;output;
function='label';color='black';size=2;style='special';text='J';x=9.73;y=5.26;output;
function='label';color='black';size=2;style='special';text='J';x=9.65;y=5.42;output;
function='label';color='black';size=2;style='special';text='J';x=9.77;y=5.48;output;
function='label';color='black';size=2;style='special';text='J';x=9.75;y=5.25;output;
function='label';color='black';size=2;style='special';text='J';x=9.82;y=5.13;output;
function='label';color='black';size=2;style='special';text='J';x=9.74;y=5.28;output;
function='label';color='black';size=2;style='special';text='J';x=9.93;y=5.31;output;
function='label';color='black';size=2;style='special';text='J';x=9.85;y=5.36;output;
function='label';color='black';size=2;style='special';text='J';x=9.91;y=5.42;output;
function='label';color='black';size=2;style='special';text='J';x=9.92;y=5.12;output;
function='label';color='black';size=2;style='special';text='J';x=9.87;y=5.45;output;
function='label';color='black';size=2;style='special';text='J';x=12.09;y=1.68;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=12.59;y=1.68;output;
function='label';color='black';size=2;style='special';text='J';x=12.91;y=1.48;output;
function='label';color='black';size=2;style='special';text='J';x=12.89;y=1.46;output;
function='label';color='black';size=2;style='special';text='J';x=12.91;y=1.85;output;
function='label';color='black';size=2;style='special';text='J';x=12.95;y=1.89;output;
function='label';color='black';size=2;style='special';text='J';x=12.96;y=1.92;output;
function='label';color='black';size=2;style='special';text='J';x=12.72;y=2.68;output;
function='label';color='black';size=2;style='special';text='J';x=12.96;y=2.59;output;
function='label';color='black';size=2;style='special';text='J';x=12.73;y=2.34;output;
function='label';color='black';size=2;style='special';text='J';x=12.73;y=2.28;output;
function='label';color='black';size=2;style='special';text='J';x=12.83;y=2.25;output;
function='label';color='black';size=2;style='special';text='J';x=12.84;y=2.19;output;
function='label';color='black';size=2;style='special';text='J';x=12.92;y=2.13;output;
function='label';color='black';size=2;style='special';text='J';x=12.85;y=2.21;output;
function='label';color='black';size=2;style='special';text='J';x=12.89;y=2.22;output;
function='label';color='black';size=2;style='special';text='J';x=12.96;y=2.22;output;
function='label';color='black';size=2;style='special';text='J';x=13.09;y=1.51;output;
function='label';color='black';size=2;style='special';text='J';x=13.09;y=1.38;output;
function='label';color='black';size=2;style='special';text='J';x=13.25;y=1.77;output;
function='label';color='black';size=2;style='special';text='J';x=13.97;y=1.57;output;
function='label';color='black';size=2;style='special';text='J';x=13.96;y=1.53;output;
function='label';color='black';size=2;style='special';text='J';x=13.97;y=1.73;output;
function='label';color='black';size=2;style='special';text='J';x=13.13;y=1.87;output;
function='label';color='black';size=2;style='special';text='J';x=13.13;y=1.91;output;
function='label';color='black';size=2;style='special';text='J';x=13.08;y=1.95;output;
function='label';color='black';size=2;style='special';text='J';x=13.09;y=1.97;output;
function='label';color='black';size=2;style='special';text='J';x=13.35;y=1.68;output;
function='label';color='black';size=2;style='special';text='J';x=13.38;y=1.78;output;
function='label';color='black';size=2;style='special';text='J';x=13.28;y=1.88;output;
function='label';color='black';size=2;style='special';text='J';x=13.29;y=1.93;output;
function='label';color='black';size=2;style='special';text='J';x=13.32;y=1.68;output;
function='label';color='black';size=2;style='special';text='J';x=13.43;y=1.78;output;
function='label';color='black';size=2;style='special';text='J';x=13.45;y=1.69;output;
function='label';color='black';size=2;style='special';text='J';x=13.45;y=1.70;output;
function='label';color='black';size=2;style='special';text='J';x=13.55;y=1.78;output;
function='label';color='black';size=2;style='special';text='J';x=13.59;y=1.82;output;
function='label';color='black';size=2;style='special';text='J';x=13.51;y=1.84;output;
function='label';color='black';size=2;style='special';text='J';x=13.54;y=1.68;output;
function='label';color='black';size=2;style='special';text='J';x=13.51;y=2.15;output;
function='label';color='black';size=2;style='special';text='J';x=13.87;y=2.61;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=13.22;y=2.62;output;
function='label';color='black';size=2;style='special';text='J';x=13.05;y=2.02;output;
function='label';color='black';size=2;style='special';text='J';x=13.07;y=2.11;output;
function='label';color='black';size=2;style='special';text='J';x=13.08;y=2.12;output;
function='label';color='black';size=2;style='special';text='J';x=13.08;y=2.09;output;
function='label';color='black';size=2;style='special';text='J';x=13.08;y=2.11;output;
run;



*Example Adaptive Cluster Sampling Steven K. Thompson (2002);
data map;
x=0;y=0;id=1;output;
x=0;y=20;id=1;output;
x=20;y=20;id=1;output;
x=20;y=0;id=1;output;
run;
goptions reset=all reset=global ftitle='Verdana';
title "Data Points";
data a;id=1000;v=1;run;
proc gmap data=a map=map all;
id id;
choro v / nolegend anno=points2;
run;
quit;
%grid(minx=0,maxx=20,miny=0,maxy=20,dim=20,anno=points2);
*%grid(minx=0,maxx=20,miny=0,maxy=20,dim=20,anno=points2,printn=NO);
data sample20;input id @@;cards;
62 76 77 86 119 207 231 301 358 372
;
%as(data=a20,sample=sample20,n=10,out=adaps,map=trab20,
 id=id,anno=points2);
%as(data=a20,sample=sample20,n=10,out=adaps,map=trab20,
 id=id,anno=points2,typen=queen);
*%as(data=a20,sample=sample20,n=10,out=adaps,map=trab20,
 id=id,anno=points2,printn=NO);
```

## B.2 Data - Example Stratified Adaptive Cluster Sampling

```
*Example Stratified Adaptive Cluster Sampling Steven K. Thompson (2002);
data points3;length text $4.;retain hsys '3' xsys ysys '2' when 'a';
/*3Pts*/
function='label';color='black';size=2;style='special';text='J';x=2.4;y=6.59;output;
function='label';color='black';size=2;style='special';text='J';x=2.8;y=6.9;output;
function='label';color='black';size=2;style='special';text='J';x=2.9;y=6.19;output;
/*16Pts*/
function='label';color='black';size=2;style='special';text='J';x=3.06;y=5.23;output;
function='label';color='black';size=2;style='special';text='J';x=3.08;y=5.31;output;
function='label';color='black';size=2;style='special';text='J';x=3.57;y=5.57;output;
function='label';color='black';size=2;style='special';text='J';x=3.65;y=5.65;output;
function='label';color='black';size=2;style='special';text='J';x=3.75;y=5.71;output;
function='label';color='black';size=2;style='special';text='J';x=3.87;y=5.74;output;
function='label';color='black';size=2;style='special';text='J';x=3.91;y=5.66;output;
function='label';color='black';size=2;style='special';text='J';x=3.95;y=5.64;output;
function='label';color='black';size=2;style='special';text='J';x=3.33;y=5.82;output;
function='label';color='black';size=2;style='special';text='J';x=3.34;y=5.87;output;
function='label';color='black';size=2;style='special';text='J';x=3.54;y=5.96;output;
function='label';color='black';size=2;style='special';text='J';x=3.21;y=5.98;output;
function='label';color='black';size=2;style='special';text='J';x=3.27;y=5.94;output;
function='label';color='black';size=2;style='special';text='J';x=3.31;y=5.89;output;
function='label';color='black';size=2;style='special';text='J';x=3.38;y=5.88;output;
function='label';color='black';size=2;style='special';text='J';x=3.43;y=5.86;output;
/*63Pts*/
function='label';color='black';size=2;style='special';text='J';x=3.51;y=6.93;output;
function='label';color='black';size=2;style='special';text='J';x=3.89;y=6.07;output;
function='label';color='black';size=2;style='special';text='J';x=3.91;y=6.08;output;
function='label';color='black';size=2;style='special';text='J';x=3.93;y=6.09;output;
function='label';color='black';size=2;style='special';text='J';x=3.94;y=6.11;output;
function='label';color='black';size=2;style='special';text='J';x=3.85;y=6.31;output;
function='label';color='black';size=2;style='special';text='J';x=3.95;y=6.29;output;
function='label';color='black';size=2;style='special';text='J';x=3.86;y=6.18;output;
function='label';color='black';size=2;style='special';text='J';x=3.92;y=6.15;output;
function='label';color='black';size=2;style='special';text='J';x=3.91;y=6.13;output;
function='label';color='black';size=2;style='special';text='J';x=3.72;y=6.79;output;
function='label';color='black';size=2;style='special';text='J';x=3.74;y=6.91;output;
function='label';color='black';size=2;style='special';text='J';x=3.86;y=6.83;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=3.88;y=6.87;output;
function='label';color='black';size=2;style='special';text='J';x=3.81;y=6.94;output;
function='label';color='black';size=2;style='special';text='J';x=3.83;y=6.95;output;
function='label';color='black';size=2;style='special';text='J';x=3.95;y=6.69;output;
function='label';color='black';size=2;style='special';text='J';x=3.97;y=6.81;output;
function='label';color='black';size=2;style='special';text='J';x=3.91;y=6.92;output;
function='label';color='black';size=2;style='special';text='J';x=3.93;y=6.96;output;
function='label';color='black';size=2;style='special';text='J';x=3.72;y=6.79;output;
function='label';color='black';size=2;style='special';text='J';x=3.74;y=6.61;output;
function='label';color='black';size=2;style='special';text='J';x=3.86;y=6.63;output;
function='label';color='black';size=2;style='special';text='J';x=3.78;y=6.87;output;
function='label';color='black';size=2;style='special';text='J';x=3.82;y=6.78;output;
function='label';color='black';size=2;style='special';text='J';x=3.94;y=6.98;output;
function='label';color='black';size=2;style='special';text='J';x=3.96;y=6.97;output;
function='label';color='black';size=2;style='special';text='J';x=3.88;y=6.94;output;
function='label';color='black';size=2;style='special';text='J';x=3.79;y=6.93;output;
function='label';color='black';size=2;style='special';text='J';x=3.78;y=6.68;output;
function='label';color='black';size=2;style='special';text='J';x=3.78;y=6.58;output;
function='label';color='black';size=2;style='special';text='J';x=3.31;y=6.61;output;
function='label';color='black';size=2;style='special';text='J';x=3.22;y=6.58;output;
function='label';color='black';size=2;style='special';text='J';x=3.28;y=6.48;output;
function='label';color='black';size=2;style='special';text='J';x=3.52;y=6.45;output;
function='label';color='black';size=2;style='special';text='J';x=3.45;y=6.43;output;
function='label';color='black';size=2;style='special';text='J';x=3.58;y=6.38;output;
function='label';color='black';size=2;style='special';text='J';x=3.61;y=6.35;output;
function='label';color='black';size=2;style='special';text='J';x=3.34;y=6.33;output;
function='label';color='black';size=2;style='special';text='J';x=3.48;y=6.28;output;
function='label';color='black';size=2;style='special';text='J';x=3.12;y=6.22;output;
function='label';color='black';size=2;style='special';text='J';x=3.44;y=6.24;output;
function='label';color='black';size=2;style='special';text='J';x=3.38;y=6.28;output;
function='label';color='black';size=2;style='special';text='J';x=3.13;y=6.31;output;
function='label';color='black';size=2;style='special';text='J';x=3.16;y=6.35;output;
function='label';color='black';size=2;style='special';text='J';x=3.22;y=6.37;output;
function='label';color='black';size=2;style='special';text='J';x=3.23;y=6.42;output;
function='label';color='black';size=2;style='special';text='J';x=3.31;y=6.45;output;
function='label';color='black';size=2;style='special';text='J';x=3.36;y=6.48;output;
function='label';color='black';size=2;style='special';text='J';x=3.49;y=6.51;output;
function='label';color='black';size=2;style='special';text='J';x=3.39;y=6.59;output;
function='label';color='black';size=2;style='special';text='J';x=3.55;y=6.67;output;
function='label';color='black';size=2;style='special';text='J';x=3.43;y=6.55;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=3.57;y=6.64;output;
function='label';color='black';size=2;style='special';text='J';x=3.08;y=6.05;output;
function='label';color='black';size=2;style='special';text='J';x=3.14;y=6.09;output;
function='label';color='black';size=2;style='special';text='J';x=3.27;y=6.13;output;
function='label';color='black';size=2;style='special';text='J';x=3.35;y=6.15;output;
function='label';color='black';size=2;style='special';text='J';x=3.42;y=6.06;output;
function='label';color='black';size=2;style='special';text='J';x=3.56;y=6.17;output;
function='label';color='black';size=2;style='special';text='J';x=3.64;y=6.08;output;
function='label';color='black';size=2;style='special';text='J';x=3.72;y=6.11;output;
function='label';color='black';size=2;style='special';text='J';x=3.86;y=6.05;output;
/*2Pts*/
function='label';color='black';size=2;style='special';text='J';x=3.61;y=7.19;output;
function='label';color='black';size=2;style='special';text='J';x=3.68;y=7.49;output;
/*3Pts*/
function='label';color='black';size=2;style='special';text='J';x=4.07;y=5.97;output;
function='label';color='black';size=2;style='special';text='J';x=4.14;y=5.98;output;
function='label';color='black';size=2;style='special';text='J';x=4.21;y=5.99;output;
/*9Pts*/
function='label';color='black';size=2;style='special';text='J';x=4.25;y=6.52;output;
function='label';color='black';size=2;style='special';text='J';x=4.32;y=6.51;output;
function='label';color='black';size=2;style='special';text='J';x=4.38;y=6.53;output;
function='label';color='black';size=2;style='special';text='J';x=4.09;y=6.49;output;
function='label';color='black';size=2;style='special';text='J';x=4.08;y=6.43;output;
function='label';color='black';size=2;style='special';text='J';x=4.11;y=6.36;output;
function='label';color='black';size=2;style='special';text='J';x=4.12;y=6.39;output;
function='label';color='black';size=2;style='special';text='J';x=4.18;y=6.36;output;
function='label';color='black';size=2;style='special';text='J';x=4.22;y=6.47;output;
/*2Pts*/
function='label';color='black';size=2;style='special';text='J';x=8.88;y=4.17;output;
function='label';color='black';size=2;style='special';text='J';x=8.91;y=4.86;output;
/*2Pts*/
function='label';color='black';size=2;style='special';text='J';x=8.78;y=5.33;output;
function='label';color='black';size=2;style='special';text='J';x=8.94;y=5.11;output;
/*5Pts*/
function='label';color='black';size=2;style='special';text='J';x=9.25;y=3.79;output;
function='label';color='black';size=2;style='special';text='J';x=9.68;y=3.76;output;
function='label';color='black';size=2;style='special';text='J';x=9.48;y=3.87;output;
function='label';color='black';size=2;style='special';text='J';x=9.52;y=3.95;output;
function='label';color='black';size=2;style='special';text='J';x=9.78;y=3.94;output;
/*57Pts*/
```

```
function='label';color='black';size=2;style='special';text='J';x=9.05;y=4.87;output;
function='label';color='black';size=2;style='special';text='J';x=9.21;y=4.86;output;
function='label';color='black';size=2;style='special';text='J';x=9.14;y=4.84;output;
function='label';color='black';size=2;style='special';text='J';x=9.26;y=4.75;output;
function='label';color='black';size=2;style='special';text='J';x=9.36;y=4.53;output;
function='label';color='black';size=2;style='special';text='J';x=9.07;y=4.56;output;
function='label';color='black';size=2;style='special';text='J';x=9.11;y=4.58;output;
function='label';color='black';size=2;style='special';text='J';x=9.46;y=4.35;output;
function='label';color='black';size=2;style='special';text='J';x=9.19;y=4.53;output;
function='label';color='black';size=2;style='special';text='J';x=9.24;y=4.42;output;
function='label';color='black';size=2;style='special';text='J';x=9.65;y=4.75;output;
function='label';color='black';size=2;style='special';text='J';x=9.48;y=4.89;output;
function='label';color='black';size=2;style='special';text='J';x=9.56;y=4.65;output;
function='label';color='black';size=2;style='special';text='J';x=9.72;y=4.66;output;
function='label';color='black';size=2;style='special';text='J';x=9.54;y=4.95;output;
function='label';color='black';size=2;style='special';text='J';x=9.61;y=4.92;output;
function='label';color='black';size=2;style='special';text='J';x=9.65;y=4.96;output;
function='label';color='black';size=2;style='special';text='J';x=9.63;y=4.89;output;
function='label';color='black';size=2;style='special';text='J';x=9.78;y=4.93;output;
function='label';color='black';size=2;style='special';text='J';x=9.81;y=4.97;output;
function='label';color='black';size=2;style='special';text='J';x=9.79;y=4.94;output;
function='label';color='black';size=2;style='special';text='J';x=9.82;y=4.96;output;
function='label';color='black';size=2;style='special';text='J';x=9.95;y=4.73;output;
function='label';color='black';size=2;style='special';text='J';x=9.91;y=4.65;output;
function='label';color='black';size=2;style='special';text='J';x=9.92;y=4.92;output;
function='label';color='black';size=2;style='special';text='J';x=9.93;y=4.89;output;
function='label';color='black';size=2;style='special';text='J';x=9.97;y=4.95;output;
function='label';color='black';size=2;style='special';text='J';x=9.94;y=4.91;output;
function='label';color='black';size=2;style='special';text='J';x=9.86;y=4.75;output;
function='label';color='black';size=2;style='special';text='J';x=9.89;y=4.92;output;
function='label';color='black';size=2;style='special';text='J';x=9.76;y=4.83;output;
function='label';color='black';size=2;style='special';text='J';x=9.83;y=4.78;output;
function='label';color='black';size=2;style='special';text='J';x=9.89;y=4.69;output;
function='label';color='black';size=2;style='special';text='J';x=9.91;y=4.62;output;
function='label';color='black';size=2;style='special';text='J';x=9.94;y=4.58;output;
function='label';color='black';size=2;style='special';text='J';x=9.86;y=4.65;output;
function='label';color='black';size=2;style='special';text='J';x=9.96;y=4.55;output;
function='label';color='black';size=2;style='special';text='J';x=9.92;y=4.48;output;
function='label';color='black';size=2;style='special';text='J';x=9.98;y=4.37;output;
function='label';color='black';size=2;style='special';text='J';x=9.69;y=4.61;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=9.32;y=4.11;output;
function='label';color='black';size=2;style='special';text='J';x=9.31;y=4.15;output;
function='label';color='black';size=2;style='special';text='J';x=9.32;y=4.25;output;
function='label';color='black';size=2;style='special';text='J';x=9.31;y=4.31;output;
function='label';color='black';size=2;style='special';text='J';x=9.48;y=4.51;output;
function='label';color='black';size=2;style='special';text='J';x=9.62;y=4.11;output;
function='label';color='black';size=2;style='special';text='J';x=9.61;y=4.15;output;
function='label';color='black';size=2;style='special';text='J';x=9.62;y=4.25;output;
function='label';color='black';size=2;style='special';text='J';x=9.61;y=4.31;output;
function='label';color='black';size=2;style='special';text='J';x=9.48;y=4.58;output;
function='label';color='black';size=2;style='special';text='J';x=9.96;y=4.11;output;
function='label';color='black';size=2;style='special';text='J';x=9.92;y=4.25;output;
function='label';color='black';size=2;style='special';text='J';x=9.98;y=4.34;output;
function='label';color='black';size=2;style='special';text='J';x=9.93;y=4.51;output;
function='label';color='black';size=2;style='special';text='J';x=9.57;y=4.63;output;
function='label';color='black';size=2;style='special';text='J';x=9.61;y=4.52;output;
function='label';color='black';size=2;style='special';text='J';x=9.42;y=4.45;output;
/*12Pts*/
function='label';color='black';size=2;style='special';text='J';x=9.12;y=5.33;output;
function='label';color='black';size=2;style='special';text='J';x=9.58;y=5.32;output;
function='label';color='black';size=2;style='special';text='J';x=9.86;y=5.44;output;
function='label';color='black';size=2;style='special';text='J';x=9.95;y=5.65;output;
function='label';color='black';size=2;style='special';text='J';x=9.96;y=5.09;output;
function='label';color='black';size=2;style='special';text='J';x=9.82;y=5.06;output;
function='label';color='black';size=2;style='special';text='J';x=9.78;y=5.08;output;
function='label';color='black';size=2;style='special';text='J';x=9.56;y=5.09;output;
function='label';color='black';size=2;style='special';text='J';x=9.79;y=5.11;output;
function='label';color='black';size=2;style='special';text='J';x=9.13;y=5.12;output;
function='label';color='black';size=2;style='special';text='J';x=9.59;y=5.12;output;
function='label';color='black';size=2;style='special';text='J';x=9.54;y=5.11;output;
/*14Pts*/
function='label';color='black';size=2;style='special';text='J';x=10.23;y=3.33;output;
function='label';color='black';size=2;style='special';text='J';x=10.48;y=3.25;output;
function='label';color='black';size=2;style='special';text='J';x=10.52;y=3.61;output;
function='label';color='black';size=2;style='special';text='J';x=10.63;y=3.62;output;
function='label';color='black';size=2;style='special';text='J';x=10.65;y=3.68;output;
function='label';color='black';size=2;style='special';text='J';x=10.89;y=3.41;output;
function='label';color='black';size=2;style='special';text='J';x=10.93;y=3.49;output;
function='label';color='black';size=2;style='special';text='J';x=10.12;y=3.98;output;
function='label';color='black';size=2;style='special';text='J';x=10.21;y=3.97;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=10.37;y=3.95;output;
function='label';color='black';size=2;style='special';text='J';x=10.51;y=3.91;output;
function='label';color='black';size=2;style='special';text='J';x=10.68;y=3.89;output;
function='label';color='black';size=2;style='special';text='J';x=10.75;y=3.92;output;
function='label';color='black';size=2;style='special';text='J';x=10.92;y=3.96;output;
/*65Pts*/
function='label';color='black';size=2;style='special';text='J';x=10.35;y=4.31;output;
function='label';color='black';size=2;style='special';text='J';x=10.31;y=4.33;output;
function='label';color='black';size=2;style='special';text='J';x=10.43;y=4.32;output;
function='label';color='black';size=2;style='special';text='J';x=10.41;y=4.12;output;
function='label';color='black';size=2;style='special';text='J';x=10.42;y=4.35;output;
function='label';color='black';size=2;style='special';text='J';x=10.32;y=4.41;output;
function='label';color='black';size=2;style='special';text='J';x=10.49;y=4.54;output;
function='label';color='black';size=2;style='special';text='J';x=10.63;y=4.09;output;
function='label';color='black';size=2;style='special';text='J';x=10.52;y=4.18;output;
function='label';color='black';size=2;style='special';text='J';x=10.59;y=4.21;output;
function='label';color='black';size=2;style='special';text='J';x=10.81;y=4.12;output;
function='label';color='black';size=2;style='special';text='J';x=10.89;y=4.25;output;
function='label';color='black';size=2;style='special';text='J';x=10.94;y=4.16;output;
function='label';color='black';size=2;style='special';text='J';x=10.97;y=4.05;output;
function='label';color='black';size=2;style='special';text='J';x=10.33;y=4.07;output;
function='label';color='black';size=2;style='special';text='J';x=10.36;y=4.11;output;
function='label';color='black';size=2;style='special';text='J';x=10.41;y=4.33;output;
function='label';color='black';size=2;style='special';text='J';x=10.55;y=4.45;output;
function='label';color='black';size=2;style='special';text='J';x=10.65;y=4.55;output;
function='label';color='black';size=2;style='special';text='J';x=10.71;y=4.65;output;
function='label';color='black';size=2;style='special';text='J';x=10.57;y=4.52;output;
function='label';color='black';size=2;style='special';text='J';x=10.78;y=4.42;output;
function='label';color='black';size=2;style='special';text='J';x=10.53;y=4.47;output;
function='label';color='black';size=2;style='special';text='J';x=10.65;y=4.49;output;
function='label';color='black';size=2;style='special';text='J';x=10.58;y=4.53;output;
function='label';color='black';size=2;style='special';text='J';x=10.35;y=4.62;output;
function='label';color='black';size=2;style='special';text='J';x=10.42;y=4.73;output;
function='label';color='black';size=2;style='special';text='J';x=10.55;y=4.65;output;
function='label';color='black';size=2;style='special';text='J';x=10.63;y=4.76;output;
function='label';color='black';size=2;style='special';text='J';x=10.67;y=4.63;output;
function='label';color='black';size=2;style='special';text='J';x=10.08;y=4.49;output;
function='label';color='black';size=2;style='special';text='J';x=10.09;y=4.65;output;
function='label';color='black';size=2;style='special';text='J';x=10.11;y=4.77;output;
function='label';color='black';size=2;style='special';text='J';x=10.15;y=4.53;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=10.17;y=4.61;output;
function='label';color='black';size=2;style='special';text='J';x=10.14;y=4.58;output;
function='label';color='black';size=2;style='special';text='J';x=10.22;y=4.89;output;
function='label';color='black';size=2;style='special';text='J';x=10.13;y=4.94;output;
function='label';color='black';size=2;style='special';text='J';x=10.26;y=4.96;output;
function='label';color='black';size=2;style='special';text='J';x=10.31;y=4.91;output;
function='label';color='black';size=2;style='special';text='J';x=10.45;y=4.96;output;
function='label';color='black';size=2;style='special';text='J';x=10.53;y=4.89;output;
function='label';color='black';size=2;style='special';text='J';x=10.54;y=4.92;output;
function='label';color='black';size=2;style='special';text='J';x=10.67;y=4.98;output;
function='label';color='black';size=2;style='special';text='J';x=10.72;y=4.91;output;
function='label';color='black';size=2;style='special';text='J';x=10.96;y=4.99;output;
function='label';color='black';size=2;style='special';text='J';x=10.97;y=4.92;output;
function='label';color='black';size=2;style='special';text='J';x=10.89;y=4.89;output;
function='label';color='black';size=2;style='special';text='J';x=10.91;y=4.93;output;
function='label';color='black';size=2;style='special';text='J';x=10.96;y=4.94;output;
function='label';color='black';size=2;style='special';text='J';x=10.93;y=4.31;output;
function='label';color='black';size=2;style='special';text='J';x=10.95;y=4.42;output;
function='label';color='black';size=2;style='special';text='J';x=10.78;y=4.90;output;
function='label';color='black';size=2;style='special';text='J';x=10.66;y=4.54;output;
function='label';color='black';size=2;style='special';text='J';x=10.38;y=4.66;output;
function='label';color='black';size=2;style='special';text='J';x=10.42;y=4.78;output;
function='label';color='black';size=2;style='special';text='J';x=10.55;y=4.58;output;
function='label';color='black';size=2;style='special';text='J';x=10.67;y=4.37;output;
function='label';color='black';size=2;style='special';text='J';x=10.82;y=4.83;output;
function='label';color='black';size=2;style='special';text='J';x=10.54;y=4.75;output;
function='label';color='black';size=2;style='special';text='J';x=10.07;y=4.54;output;
function='label';color='black';size=2;style='special';text='J';x=10.24;y=4.95;output;
function='label';color='black';size=2;style='special';text='J';x=10.35;y=4.97;output;
function='label';color='black';size=2;style='special';text='J';x=10.96;y=4.73;output;
function='label';color='black';size=2;style='special';text='J';x=10.83;y=4.65;output;
/*12Pts*/
function='label';color='black';size=2;style='special';text='J';x=10.59;y=5.58;output;
function='label';color='black';size=2;style='special';text='J';x=10.57;y=5.56;output;
function='label';color='black';size=2;style='special';text='J';x=10.08;y=5.09;output;
function='label';color='black';size=2;style='special';text='J';x=10.13;y=5.16;output;
function='label';color='black';size=2;style='special';text='J';x=10.27;y=5.08;output;
function='label';color='black';size=2;style='special';text='J';x=10.35;y=5.17;output;
function='label';color='black';size=2;style='special';text='J';x=10.45;y=5.14;output;
function='label';color='black';size=2;style='special';text='J';x=10.44;y=5.11;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=10.56;y=5.13;output;
function='label';color='black';size=2;style='special';text='J';x=10.63;y=5.15;output;
function='label';color='black';size=2;style='special';text='J';x=10.72;y=5.09;output;
function='label';color='black';size=2;style='special';text='J';x=10.89;y=5.12;output;
/*5Pts*/
function='label';color='black';size=2;style='special';text='J';x=11.65;y=3.85;output;
function='label';color='black';size=2;style='special';text='J';x=11.07;y=3.96;output;
function='label';color='black';size=2;style='special';text='J';x=11.08;y=3.58;output;
function='label';color='black';size=2;style='special';text='J';x=11.11;y=3.89;output;
function='label';color='black';size=2;style='special';text='J';x=11.21;y=3.98;output;
/*17 Pts*/
function='label';color='black';size=2;style='special';text='J';x=11.95;y=4.87;output;
function='label';color='black';size=2;style='special';text='J';x=11.59;y=4.95;output;
function='label';color='black';size=2;style='special';text='J';x=11.46;y=4.87;output;
function='label';color='black';size=2;style='special';text='J';x=11.45;y=4.67;output;
function='label';color='black';size=2;style='special';text='J';x=11.26;y=4.88;output;
function='label';color='black';size=2;style='special';text='J';x=11.25;y=4.68;output;
function='label';color='black';size=2;style='special';text='J';x=11.23;y=4.85;output;
function='label';color='black';size=2;style='special';text='J';x=11.18;y=4.88;output;
function='label';color='black';size=2;style='special';text='J';x=11.07;y=4.97;output;
function='label';color='black';size=2;style='special';text='J';x=11.16;y=4.56;output;
function='label';color='black';size=2;style='special';text='J';x=11.05;y=4.31;output;
function='label';color='black';size=2;style='special';text='J';x=11.14;y=4.45;output;
function='label';color='black';size=2;style='special';text='J';x=11.08;y=4.78;output;
function='label';color='black';size=2;style='special';text='J';x=11.11;y=4.31;output;
function='label';color='black';size=2;style='special';text='J';x=11.07;y=4.25;output;
function='label';color='black';size=2;style='special';text='J';x=11.32;y=4.16;output;
function='label';color='black';size=2;style='special';text='J';x=11.31;y=4.21;output;
/*1Pt*/
function='label';color='black';size=2;style='special';text='J';x=11.09;y=5.42;output;
/*2Pts*/
function='label';color='black';size=2;style='special';text='J';x=16.89;y=16.76;output;
function='label';color='black';size=2;style='special';text='J';x=16.87;y=16.96;output;
/*36Pts*/
function='label';color='black';size=2;style='special';text='J';x=16.13;y=17.98;output;
function='label';color='black';size=2;style='special';text='J';x=16.06;y=17.96;output;
function='label';color='black';size=2;style='special';text='J';x=16.29;y=17.97;output;
function='label';color='black';size=2;style='special';text='J';x=16.14;y=17.99;output;
function='label';color='black';size=2;style='special';text='J';x=16.12;y=17.89;output;
function='label';color='black';size=2;style='special';text='J';x=16.31;y=17.76;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=16.41;y=17.65;output;
function='label';color='black';size=2;style='special';text='J';x=16.52;y=17.67;output;
function='label';color='black';size=2;style='special';text='J';x=16.95;y=17.89;output;
function='label';color='black';size=2;style='special';text='J';x=16.89;y=17.35;output;
function='label';color='black';size=2;style='special';text='J';x=16.49;y=17.97;output;
function='label';color='black';size=2;style='special';text='J';x=16.55;y=17.87;output;
function='label';color='black';size=2;style='special';text='J';x=16.63;y=17.98;output;
function='label';color='black';size=2;style='special';text='J';x=16.73;y=17.78;output;
function='label';color='black';size=2;style='special';text='J';x=16.91;y=17.82;output;
function='label';color='black';size=2;style='special';text='J';x=16.88;y=17.91;output;
function='label';color='black';size=2;style='special';text='J';x=16.89;y=17.89;output;
function='label';color='black';size=2;style='special';text='J';x=16.92;y=17.92;output;
function='label';color='black';size=2;style='special';text='J';x=16.95;y=17.93;output;
function='label';color='black';size=2;style='special';text='J';x=16.78;y=17.88;output;
function='label';color='black';size=2;style='special';text='J';x=16.55;y=17.09;output;
function='label';color='black';size=2;style='special';text='J';x=16.56;y=17.11;output;
function='label';color='black';size=2;style='special';text='J';x=16.59;y=17.21;output;
function='label';color='black';size=2;style='special';text='J';x=16.62;y=17.28;output;
function='label';color='black';size=2;style='special';text='J';x=16.65;y=17.31;output;
function='label';color='black';size=2;style='special';text='J';x=16.68;y=17.13;output;
function='label';color='black';size=2;style='special';text='J';x=16.71;y=17.08;output;
function='label';color='black';size=2;style='special';text='J';x=16.71;y=17.11;output;
function='label';color='black';size=2;style='special';text='J';x=16.89;y=17.23;output;
function='label';color='black';size=2;style='special';text='J';x=16.98;y=17.31;output;
function='label';color='black';size=2;style='special';text='J';x=16.83;y=17.37;output;
function='label';color='black';size=2;style='special';text='J';x=16.89;y=17.26;output;
function='label';color='black';size=2;style='special';text='J';x=16.94;y=17.38;output;
function='label';color='black';size=2;style='special';text='J';x=16.88;y=17.49;output;
function='label';color='black';size=2;style='special';text='J';x=16.85;y=17.63;output;
function='label';color='black';size=2;style='special';text='J';x=16.97;y=17.82;output;
/*23Pts*/
function='label';color='black';size=2;style='special';text='J';x=16.52;y=18.89;output;
function='label';color='black';size=2;style='special';text='J';x=16.55;y=18.86;output;
function='label';color='black';size=2;style='special';text='J';x=16.11;y=18.35;output;
function='label';color='black';size=2;style='special';text='J';x=16.19;y=18.09;output;
function='label';color='black';size=2;style='special';text='J';x=16.09;y=18.18;output;
function='label';color='black';size=2;style='special';text='J';x=16.48;y=18.37;output;
function='label';color='black';size=2;style='special';text='J';x=16.63;y=18.38;output;
function='label';color='black';size=2;style='special';text='J';x=16.92;y=18.58;output;
function='label';color='black';size=2;style='special';text='J';x=16.76;y=18.61;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=16.81;y=18.08;output;
function='label';color='black';size=2;style='special';text='J';x=16.76;y=18.31;output;
function='label';color='black';size=2;style='special';text='J';x=16.72;y=18.28;output;
function='label';color='black';size=2;style='special';text='J';x=16.66;y=18.15;output;
function='label';color='black';size=2;style='special';text='J';x=16.71;y=18.18;output;
function='label';color='black';size=2;style='special';text='J';x=16.83;y=18.39;output;
function='label';color='black';size=2;style='special';text='J';x=16.89;y=18.52;output;
function='label';color='black';size=2;style='special';text='J';x=16.96;y=18.56;output;
function='label';color='black';size=2;style='special';text='J';x=16.97;y=18.48;output;
function='label';color='black';size=2;style='special';text='J';x=16.98;y=18.62;output;
function='label';color='black';size=2;style='special';text='J';x=16.92;y=18.49;output;
function='label';color='black';size=2;style='special';text='J';x=16.89;y=18.11;output;
function='label';color='black';size=2;style='special';text='J';x=16.72;y=18.18;output;
function='label';color='black';size=2;style='special';text='J';x=16.96;y=18.26;output;
/*34 Pts*/
function='label';color='black';size=2;style='special';text='J';x=17.21;y=17.06;output;
function='label';color='black';size=2;style='special';text='J';x=17.22;y=17.13;output;
function='label';color='black';size=2;style='special';text='J';x=17.32;y=17.18;output;
function='label';color='black';size=2;style='special';text='J';x=17.41;y=17.27;output;
function='label';color='black';size=2;style='special';text='J';x=17.79;y=17.31;output;
function='label';color='black';size=2;style='special';text='J';x=17.65;y=17.51;output;
function='label';color='black';size=2;style='special';text='J';x=17.66;y=17.55;output;
function='label';color='black';size=2;style='special';text='J';x=17.79;y=17.78;output;
function='label';color='black';size=2;style='special';text='J';x=17.78;y=17.81;output;
function='label';color='black';size=2;style='special';text='J';x=17.77;y=17.79;output;
function='label';color='black';size=2;style='special';text='J';x=17.08;y=17.48;output;
function='label';color='black';size=2;style='special';text='J';x=17.09;y=17.51;output;
function='label';color='black';size=2;style='special';text='J';x=17.16;y=17.63;output;
function='label';color='black';size=2;style='special';text='J';x=17.11;y=17.61;output;
function='label';color='black';size=2;style='special';text='J';x=17.25;y=17.75;output;
function='label';color='black';size=2;style='special';text='J';x=17.23;y=17.45;output;
function='label';color='black';size=2;style='special';text='J';x=17.38;y=17.55;output;
function='label';color='black';size=2;style='special';text='J';x=17.32;y=17.83;output;
function='label';color='black';size=2;style='special';text='J';x=17.45;y=17.92;output;
function='label';color='black';size=2;style='special';text='J';x=17.48;y=17.85;output;
function='label';color='black';size=2;style='special';text='J';x=17.09;y=17.71;output;
function='label';color='black';size=2;style='special';text='J';x=17.11;y=17.83;output;
function='label';color='black';size=2;style='special';text='J';x=17.25;y=17.96;output;
function='label';color='black';size=2;style='special';text='J';x=17.37;y=17.98;output;
function='label';color='black';size=2;style='special';text='J';x=17.48;y=17.99;output;
```

```
function='label';color='black';size=2;style='special';text='J';x=17.59;y=17.97;output;
function='label';color='black';size=2;style='special';text='J';x=17.65;y=17.98;output;
function='label';color='black';size=2;style='special';text='J';x=17.77;y=17.89;output;
function='label';color='black';size=2;style='special';text='J';x=17.78;y=17.97;output;
function='label';color='black';size=2;style='special';text='J';x=17.04;y=17.99;output;
function='label';color='black';size=2;style='special';text='J';x=17.05;y=17.55;output;
function='label';color='black';size=2;style='special';text='J';x=17.02;y=17.61;output;
function='label';color='black';size=2;style='special';text='J';x=17.07;y=17.73;output;
function='label';color='black';size=2;style='special';text='J';x=17.11;y=17.82;output;
/*14Pts*/
function='label';color='black';size=2;style='special';text='J';x=17.45;y=18.98;output;
function='label';color='black';size=2;style='special';text='J';x=17.44;y=18.97;output;
function='label';color='black';size=2;style='special';text='J';x=17.16;y=18.87;output;
function='label';color='black';size=2;style='special';text='J';x=17.17;y=18.88;output;
function='label';color='black';size=2;style='special';text='J';x=17.09;y=18.76;output;
function='label';color='black';size=2;style='special';text='J';x=17.11;y=18.62;output;
function='label';color='black';size=2;style='special';text='J';x=17.24;y=18.55;output;
function='label';color='black';size=2;style='special';text='J';x=17.13;y=18.38;output;
function='label';color='black';size=2;style='special';text='J';x=17.03;y=18.16;output;
function='label';color='black';size=2;style='special';text='J';x=17.21;y=18.24;output;
function='label';color='black';size=2;style='special';text='J';x=17.34;y=18.21;output;
function='label';color='black';size=2;style='special';text='J';x=17.43;y=18.18;output;
function='label';color='black';size=2;style='special';text='J';x=17.53;y=18.23;output;
function='label';color='black';size=2;style='special';text='J';x=17.57;y=18.26;output;
run;

data map;
x=0;y=0;id=1;output;
x=0;y=20;id=1;output;
x=20;y=20;id=1;output;
x=20;y=0;id=1;output;
run;
goptions reset=all reset=global ftitle='Verdana';
title "Data Points";
data a;id=1000;v=1;run;
proc gmap data=a map=map all;
id id;
choro v / nolegend anno=points3;
run;
quit;
```

```
%grid(minx=0,maxx=20,miny=0,maxy=20,dim=20,anno=points3);
data A20s;set A20;
if id<=200 then strata=1;else strata=2;
run;
data points3_;set points3 coor;run;
title "Strata";
proc gmap data=a20s map=trab20 all;
id id;
choro strata / anno=points3_;
run;
quit;
data sample20s;input id strata @@;cards;
87 1 91 1 138 1 153 1 166 1 221 2 275 2 291 2 295 2 363 2
;
%as(data=A20s,sample=sample20s,n=10,out=adaps,strata=strata,map=trab20,
 id=id,anno=points3);
%as(data=A20s,sample=sample20s,n=10,out=adaps,strata=strata,map=trab20,
 id=id,anno=points3,typen=queen);
```