AMOSTRAGEM ESPACIAL:

Contribuições metodológicas - Uma análise exploratória.

Nome: Elisete de Assis Rebello Leite Ribeiro

Dissertação de Mestrado

Orientador: Dr. Leônidas C. Barroso

Co-orientador: Dr. João Francisco de Abreu

1º Semestre/2000

Elisete de Assis Rebello Leite Ribeiro

AMOSTRAGEM ESPACIAL:
Contribuições metodológicas - Uma análise exploratória.

Dissertação apresentada ao Programa de Pós-graduação em Tratamento da Informação Espacial, como requisito parcial à obtenção do Título de Mestre.

Área de Concentração: Análise Espacial

Orientador: Prof. Dr.Leônidas Conceição Barroso

Co-orientador: Prof. Dr. João Francisco de Abreu

PUC-MG
Belo Horizonte
2000

Título: AMOSTRAGEM ESPACIAL: Contribuições Metodológicas - Uma análise exploratória.

Autor: Elisete de Assis Rebello Leite Ribeiro

Aprovada em:

Examinadores:

Programa de Pós-graduação em Tratamento da Informação Espacial
Pontifícia Universidade Católica de Minas Gerais

Aos seres irracionais, que se portam como racionais; e que nos múltiplos momentos de caminhadas, com um olhar desavisado, ficam apenas a observar....

FICHA CATALOGRÁFICA

R484a

Ribeiro, Elisete de Assis Rebello Leite
Amostragem espacial: contribuições metodológicas: uma análise exploratória / Elisete de Assis Rebello Leite Ribeiro. - Belo Horizonte, 2000.
186 f.: il.

Orientador: Prof. Dr. Leônidas Conceição Barroso.
Co-orientador: Prof. Dr. João Francisco de Abreu.
Dissertação (mestrado) - Pontifícia Universidade Católica de Minas Gerais.
Bibliografia.

1.Amostragem (Estatística). 2.Análise espacial (Estatística). 3.Ciência política - Estatística. 4. Sistemas de informação geográfica. I. Barroso , Leônidas Conceição. II. Pontifícia Universidade Católica de Minas Gerais. III. Título.

CDU: 519.2

AGRADECIMENTOS

Aos Professores, Luiz G. de Azevedo, Yara C. Nogueira e Karla F. Cipreste, pelo apoio linguístico e amizade constantes.

À professora Ionne Grossi, pela visão irrestrita.

Aos colegas e funcionários do Departamento de Matemática e Estatística, pelos conselhos, carinho e colaboração.

Ao professor Tarcísio B. de Andrade, pela atenção e ajuda dispensada a todos.

Às bibliotecárias da PUC-MG pelo atendimento e solicitude.

À Bertha Maakaroun, da Pólis Pesquisa e Consultoria, pela divisão de conhecimentos.

À PUC-MG, pelo esforço em qualificar seus profissionais.

À Prefeitura de Ibirité, pela cessão do traçado urbano do município.

Ao professor José Francisco Soares, pela confiança.

Ao professor Leônidas C. Barroso, pela paciência e dedicação.

Ao professor João Francisco de Abreu, pela oportunidade dada frente aos desafios impostos pelo trabalho.

À minha família, pela espera interminável.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

A.a.	Amostra Aleatória
CEFET	Centro Federal de Educação Tecnológica
CPF	Correção das Populações Finitas
FAPEMIG	Fundação de Amparo à Pesquisa do Estado de Minas Gerais
GIS = SIG	Sistemas de Informações Geográficas
IBGE	Fundação Instituto Brasileiro de Geografia e Estatística
IC	Intervalo de Confiança
IGA	Instituto de Geociências Aplicadas
% Vál.	Percentual de Votos Válidos
% Comp.	Percentual de Votos Válidos Com Relação ao Comparecimento
PPT	Probabilidade Proporcional ao Tamanho
Qtd. ou Qtde.	Quantidade
RMBH	Região Metropolitana de Belo Horizonte
SEBRAE/MG	Serviço de Apoio às Micro e Pequenas Empresas de Minas Gerais
TCL	Teorema Central do Limite
TNA	Tabela ou Tábua de Números Aleatórios
TRE	Tribunal Regional Eleitoral
V.a.	Variável Aleatória

RESUMO

O objetivo deste trabalho é fazer um estudo sobre amostragem espacial, partindo das amostragens clássicas de probabilidades, bem como mostrar a importância das seleções não probabilísticas quando se trata de opinião de eleitores.

Para ilustrar o conteúdo teórico, mostra-se, por meio de um estudo de caso para a cidade de Ibirité, uma seleção de amostragem combinada, na qual se discute não apenas a validade da seleção de amostragem, mas erro, dimensionamento, além da espacialização das variáveis.

Essas análises serão importantes, a partir do momento em que se puder afirmar que uma seleção por amostragem tem menor variabilidade, ou que a seleção conduz a uma amostra que realmente pertença à população em estudo, isto é, que satisfaça não apenas a uma especialidade.

O entendimento da distribuição da população é de fundamental importância para se fazer as análises propostas, na falta desse conhecimento, suposições serão apresentadas durante o trabalho.

O estudo aqui mencionado, não pretende, em nenhum momento, abranger a totalidade. Buscam-se idéias que propiciem um melhor entendimento de alguns projetos de amostragem e no estudo de caso, o porquê da melhor qualidade entre alguns projetos. Entretanto, o importante é o projeto amostral em si mesmo, isto é, os diversos motivos capazes de conduzir ao projeto de amostragem "ideal", que conduza a resultados satisfatórios e amplie o conhecimento para melhores seleções amostrais no futuro.

1. INTRODUÇÃO

1.1 APRESENTAÇÃO

A seleção de amostragem é de extrema importância para qualquer estudo em campo, principalmente quanto à espacialização dos fenômenos. O estudo aqui desenvolvido parte das amostragens clássicas e chega às seleções espacializadas, associando o espaço às seleções de amostragem.

O espaço nem sempre é trabalhado conjuntamente com as seleções de amostragem, e isto se deve à falta de informações de maior complexidade sobre a distribuição do fenômeno no próprio espaço. Desta forma, ao se fazer uma seleção de amostragem, nem sempre o espaço é considerado o ente principal da seleção, e acaba por ser tratado secundariamente. As seleções adaptadoras, inseridas ao estudo (anexo 07), representam uma inovação de conceitos e aplicações diretamente relacionados ao espaço.

As seleções de amostragem não probabilísticas também foram contempladas no estudo, pois em muitos casos, devido ao próprio direcionamento deste, elas podem ser utilizadas com o benefício de fornecerem melhores resultados que as seleções probabilísticas (Adler, 1971).

Para ilustrar este trabalho, um estudo de caso foi proposto para a cidade de Ibirité. No estudo de caso, fica clara a importância da espacialização ou georeferenciamento da unidade amostrada (eleitor), como também a importância do estudo das subpopulações individualizadas e do dimensionamento da amostra.

O espaço nem sempre é trabalhado da maneira ideal, mas também não deve ser colocado como um fenômeno secundário. Quando se conhece as características e especificidades das variáveis em estudo, a seleção de amostragem e a análise de dados se tornam mais seguras, fornecendo aplicações mais refinadas, com menor variabilidade de resultados, possibilitando a utilização de seleções amostrais mais complexas como as adaptadoras (Cochran, 1997).

Para o estudo de caso proposto, a cidade de Ibirité foi dividida em regiões eleitorais (agrupamento de zonas eleitorais) que representavam estratos, e as regiões eleitorais foram subdivididas em áreas (pequenos conglomerados). Desta forma, as bases das seleções estratificadas (quotas), conglomerados (por área - espacial) e sistemática (para a coleta) foram utilizadas. Assim, o espaço não assumiu uma importância secundária, visto que a distribuição das unidades a serem amostradas (eleitores) sobre o espaço foi expandida e as áreas foram quase que completamente trabalhadas, quando da distribuição das quotas por entrevistador. As áreas foram mapeadas, o que favoreceu a distribuição da amostra sobre as regiões eleitorais.

Algumas sugestões, como o tratamento individualizado por região eleitoral, foram colocadas, pois permitem melhores resultados, apesar de terem um custo maior. Uma forma de espacialização do eleitor também foi discutida.

Em síntese, o estudo de caso coloca as seleções de amostragem, o cálculo de erro, o dimensionamento, as inferências estatísticas e, principalmente, a importância do georeferenciamento numa seleção de amostragem.

1.2 JUSTIFICATIVA / OBJETIVOS

A amostragem espacial constitui uma importante ferramenta na coleta de dados para o planejamento de uma pesquisa em campo. Com a evolução da Geografia e das tecnologias, as seleções de amostragem associadas ao espaço, na atualidade, representam a melhor forma de coleta de dados para análise de fenômenos espacializados. Principalmente quando o levantamento de dados é inviável, devido a restrições de custo, tempo, mão de obra e outros (Gerardi & Silva, 1981).

Em nível acadêmico, buscam-se condições de melhoria de entendimento das seleções de amostragem espaciais, que a partir das três últimas décadas alcançaram grande desenvolvimento teórico com importantes aplicações (Cochran, 1977).

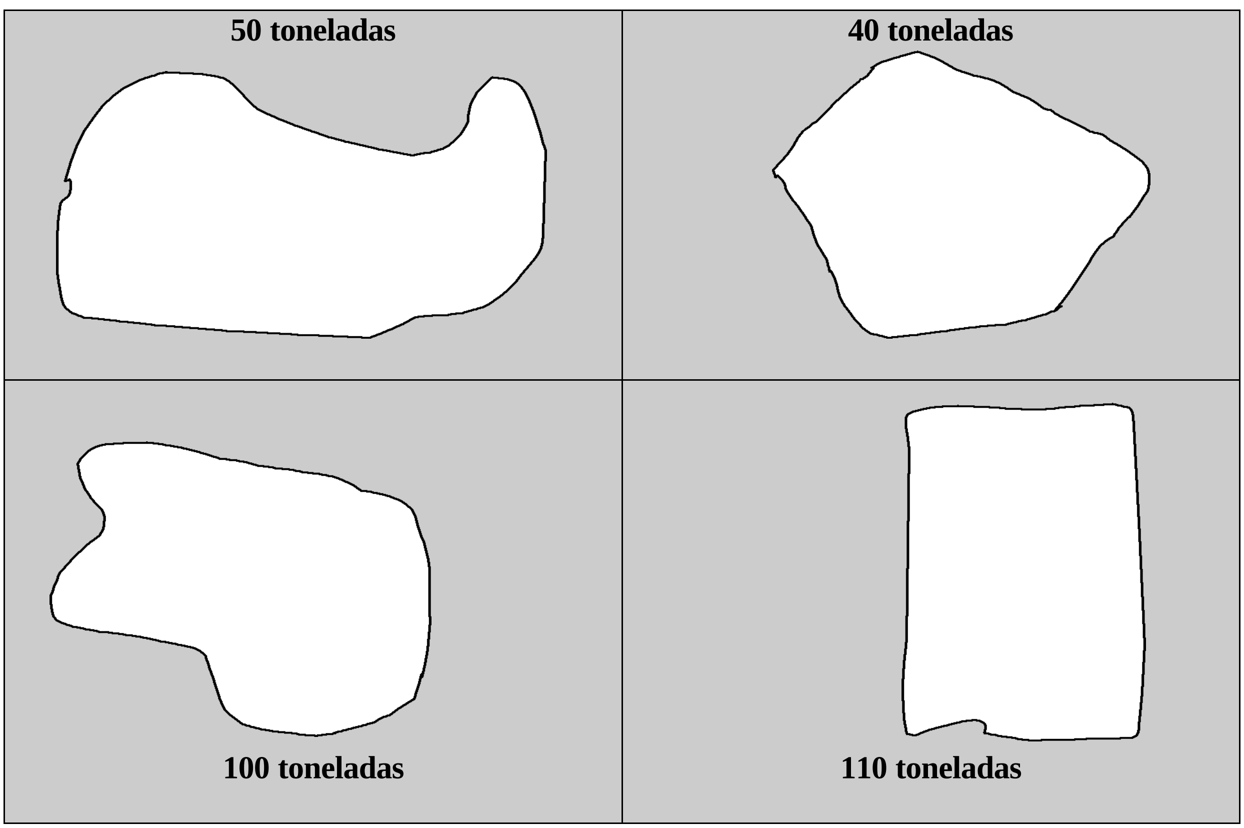
> "A amostragem tem sido usada de longa data pelos geógrafos, embora de forma não consciente e em sentido diferente do uso atual em pesquisa, e esta importante diferença é aquela que distingue a amostragem proposital, intencional, subjetiva ou não probabilística da amostragem objetiva ou probabilística" (Gerardi & Silva, 1981).

As técnicas e análises quantitativas na Geografia são muito importantes, mas devem ser utilizadas com cuidado, devido ao enfoque espacial das variáveis. Às vezes, a simplificação é tão acentuada, que o resultado pode representar uma subestimação ou superestimação da variável. Assim, tratar com subespaços georeferenciados, em muitos casos, talvez seja a melhor opção.

> ".... é preciso reconhecer que a Geografia, por aceitar durante muito tempo os seus fenômenos como sendo excepcionais, - como decorrência de sua localização individualizada sobre o espaço, não sendo por isso mesmo passíveis de sujeição a leis e princípios gerais - , utilizou e utiliza ainda hoje, muito mais e freqüentemente de forma limitada, a abordagem indutiva" (Gerardi $ Silva, 1981).

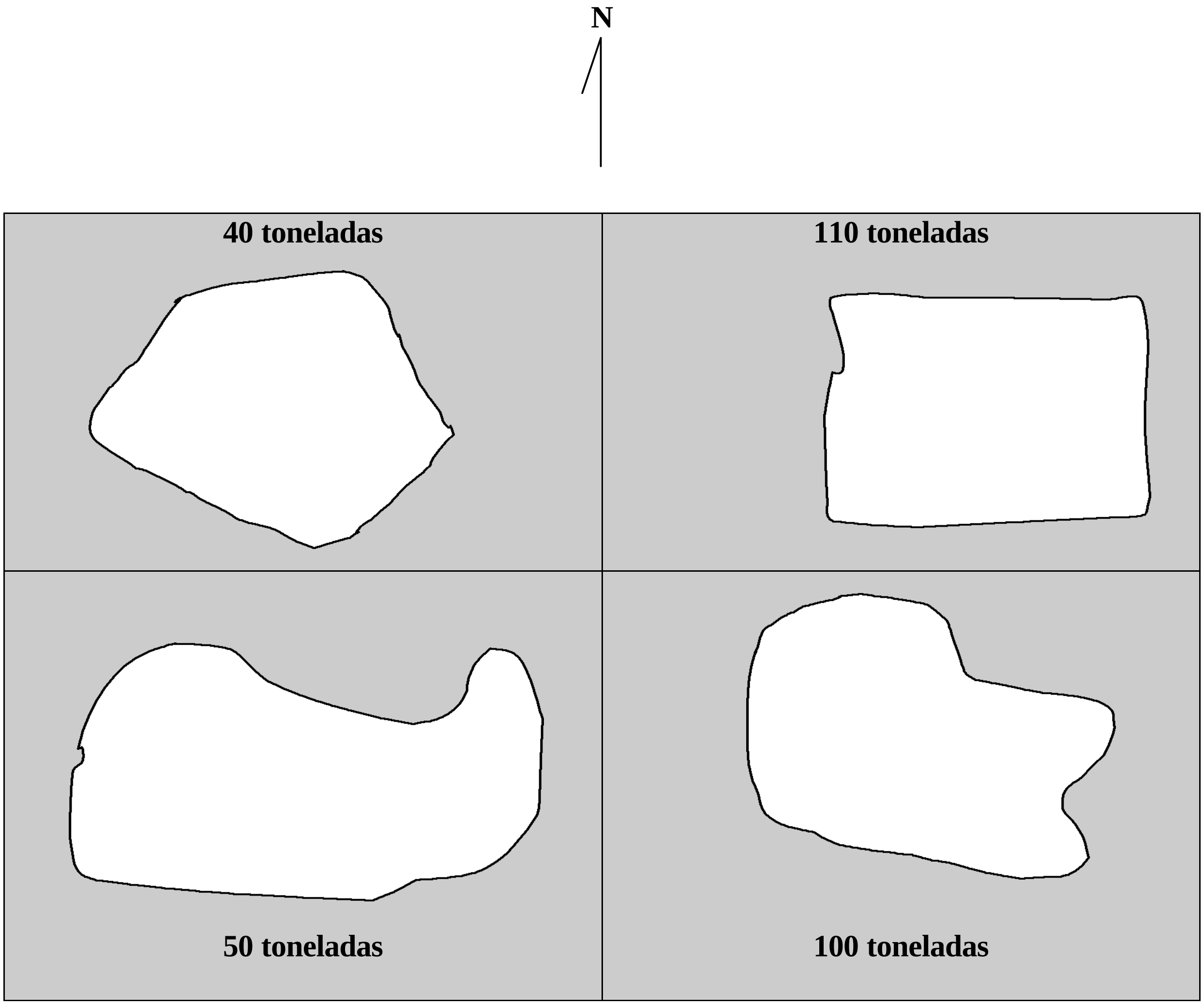
Para exemplificar a importância da análise de resultados, por meio de medidas, vamos verificar o que pode ocorrer quando se trabalha com a média de um conjunto de dados. Observe a figura 01 e 02.

FIGURA 01: Produção de Soja na Fazenda X.



Fonte: Dados hipotéticos.

FIGURA 02: Produção de Soja na Fazenda X por Área (dados georeferenciados) .



Fonte: Dados hipotéticos.

Tanto na figura 01 como na figura 02, a produção total é de 300 toneladas e a média aritmética é de 75 toneladas, isto é, as medidas são similares.

Supondo que um novo adubo (600 kg) será utilizado para melhorar a produção, e que o gerente desta fazenda tenha resolvido dividir os 600 kg em quatro partes similares, isto é, 150 kg/área. Esta seria a pior das opções, pois como se pode observar, quando se tem os dados espacializados, pode-se atuar individualmente por área. Assim, as áreas de menor produção e de maior tamanho devem receber maior quantidade de adubo. Ao se estimar quantidade, a localização tem que ser exata, pois no caso da soja, fica mais fácil visualizar o fenômeno, em contato constante com os plantadores; mas quando se trata de extração mineral, por exemplo, se os dados não estiverem georeferenciados, a extração pode se iniciar justamente de um local em que a quantidade de minério seja mínima (para isto, basta inverter o sistema de coordenadas), iniciando-se assim um processo de grandes prejuízos.

As distribuições espaciais são de interesse direto do geógrafo, assim é necessário o levantamento de dados para análise da variabilidade do fenômeno no espaço, por meio de uma seleção de amostragem.

Na Estatística espacial, as estimativas para a população podem ser únicas, mas com um ajuste no cálculo das estatísticas clássicas, nas quais a localização também faz parte do cálculo da medida. Assim, a média aritmética pode ser associada ao centro médio, a mediana ao centro mediano, a moda ao centro modal e o desvio-padrão à distância padrão (raio padrão ou raio dinâmico). Para o cálculo das medidas espaciais, levam-se em conta pares ordenados (X,Y), associados à localização das informações no espaço. As definições destas medidas se encontram no glossário.

Após a Segunda Guerra Mundial, devido às novas tecnologias, houve uma explosão de informações, tanto para as pesquisas científicas, quanto para um público mais amplo, como também para os políticos. Planejar com uma grande quantidade de dados tornou-se essencial nos dias atuais e permite uma visão mais ampla dos mecanismos e estruturas dos processos sociais, além disso, evita os altos custos e riscos de decisões erradas e as informações podem ser usadas entre as instituições como poderoso instrumento nas negociações. A análise espacial tornou-se de fundamental importância, pois o planejamento estratégico em qualquer nível passa a depender diretamente da análise espacial (Abreu, 1995).

Desta forma, este trabalho tem como objetivos:

i) Mostrar, por meio do estudo das seleções de amostragem, a importância da espacialização dos fenômenos;

ii) Apresentar as seleções clássicas de probabilidades como a base para as seleções espaciais, nas quais a teoria clássica pode ser utilizada ou estendida aos fenômenos espaciais;

iii) Colocar a importância do conhecimento do espaço como primordial na caracterização de um fenômeno, utilizando-se um estudo de caso;

iv) Sugerir, na falta de condições ideais para a seleção de amostragem, meios de melhoria ou de ajuste para a seleção;

v) Fazer um breve relato sobre dimensionamento e erro, que estão diretamente associados à precisão dos resultados (confiabilidade dos resultados).

1.3 REVISÃO BIBLIOGRÁFICA

ABREU, João Francisco. **Manufatura Integrada Por Computador; contexto , tendências e técnicas**. Belo Horizonte: Editora Terra, L.D. CNPQ / SEBRAE/ CEFET/ Minas Gerais, 1995.

Este trabalho mostra a importância dos Sistemas de Informações Geográficas (GIS) no tratamento de dados georeferenciados, em que a comparação de dados originários de diversas fontes se torna viável, evitando-se altos custos de ações não compatíveis com a realidade. O autor fornece conceitos, estruturas e principais usos desses sistemas e desenvolve um breve relato histórico e técnico-científico da evolução dos mesmos.

BURT, James E. ; BARBER , Gerald M.. **Elementaty Statistics for Geographers**. 2 ed. New York: The Guilford Press, 1996, p. 242/251.

A amostragem geográfica com a utilização do sistema cartesiano (espacialização) é tratada neste trabalho de forma prática. A associação das amostragens clássicas às geográficas é de suma importância para o entendimento da própria seleção amostral, quanto para a utilização das estatísticas e estimativas, que são as mesmas das amostragens clássicas, quando o procedimento é para a seleção simples de amostragem. As amostragens por pontos, áreas e linhas são trabalhadas espacialmente, isto é, a localização do fenômeno é primordial para ações e melhorias. O fenômeno a ser estudado geralmente já está mapeado, isto é, as variáveis em estudo encontram-se previamente localizadas.

A inferência estatística é tratada de maneira simples, inclusive com a utilização da distribuição normal, o que favorece o entendimento.

GERARDI, Lúcia H. de Oliveira, SILVA, Bárbara Christine Maria Nentwig. **Quatificação em Geografia**. São Paulo: Difel, 1981.

É um livro para geógrafos, no qual as aplicações básicas quantitativas e estatísticas são colocadas em linguagem geográfica, para melhor entendimento. Representa um bom início para o entendimento das amostragens espaciais. Este trabalho mostra a importância da estatística, como ferramenta, auxiliar à análise e compreensão de comportamentos.

Muitas das estatísticas fazem parte desse estudo, no qual as clássicas (média, desvio-padrão, mediana, moda) estão diretamente associadas às estatísticas espaciais. O conceito de normalidade também é utilizado nesse estudo, devido ao fato de que muitas variáveis, em Geografia, assumem a condição de uma variável normal. As autoras deste trabalho também exemplificam o ajuste linear por meio da análise de correlação, residual e limites de confiança (intervalos de confiança).

TEIXEIRA, Amândio Luís de Almeida; CHRISTOFOLETTI, Antônio. **Sistemas de Informações Geográficas:** Dicionário Ilustrado. São Paulo: Hucitec, 1997.

Os autores, por meio deste dicionário, caracterizam grande parte das terminologias geográficas associadas a sistemas computacionais complexos. Na atualidade, com a evolução das técnicas associadas à computação, a leitura de trabalhos torna-se árida e difícil se o leitor não tiver domínio da linguagem, este dicionário dirime este problema.

COCHRAN, William Gemmmell. **Sampling Techniques**. New York: John Wiley & Sons, 1977.

O estudo contempla as variáveis média, proporção, índice populacional e total populacional para as amostragens clássicas (simples, estratificada, conglomerado), além de outros tipos de seleção aleatória e para a variância populacional, que são de importância relevante para o entendimento teórico das seleções amostrais. A discussão da variabilidade é constante em todo trabalho e isto pode ser avaliado nas exemplificações, nas quais se analisa a variância entre as seleções amostrais. Teoricamente, este trabalho pode ser visto como a base das seleções amostrais adaptadoras, nas quais o espaço geográfico começa a ser trabalhado e visto como de importância fundamental para as seleções amostrais.

THOMPSON, K. Steven. **Sampling**. John Wiley & Sons, Inc. New York, 1992, p.212/302.

Este livro desenvolve os conceitos de seleção de amostragem, partindo das amostragens clássicas, até chegar aos tipos de seleção adaptadoras. Para as seleções adaptadoras, ampliam-se os conceitos de seleção amostral, mostrando as diversas diferenças e aplicações. A forma de apresentação dos exemplos, para as seleções clássicas, pode ser associada diretamente à amostragem geográfica. Quanto à inferência estatística, ela é colocada de forma direta, ou seja, para se entender a matemática e a essência das estimativas é necessário entendimento anterior de estatística básica e cálculo.

As amostragens adaptadoras (simples, estratificada, conglomerados e sistemática) são trabalhadas, primeiramente, com um enfoque clássico espacial (áreas), depois são colocadas como projetos adaptadores, nos quais as estimativas são de difícil entendimento para os não especialistas.

TRIOLA, Mário F.. **Introdução À Estatística**. Rio de Janeiro: LTC, 1999.

Este trabalho reune toda a base teórica para o entendimento da estatística básica, trazendo exemplificações reais e diferenciadas que despertam a atenção do leitor. A associação entre a teoria e a prática é feita por meio do "software Minitab", com a informação dos comandos a serem utilizados. Por todo o trabalho, observam-se críticas informativas sobre o uso das estatísticas como, também, o relato histórico da evolução da estatística. É um trabalho diferente dos demais, associando teoria e prática com o auxílio do computador, o que, na atualidade, representa um avanço no entendimento da estatística.

WERKEMA, Maria Cristina Catarino. **Como Estabelecer Conclusões Com Confiança: Entendendo Inferência Estatística**. Belo Horizonte: Fundação Christiano Ottoni, 1996.

Dirigido à área de ciências exatas, exemplifica muito bem a aplicação de intervalos de confiança e teste de hipóteses para os parâmetros populacionais. Descreve sobre os estimadores de forma sucinta e mostra os intervalos de confiança individualmente para cada parâmetro populacional. Os testes de hipóteses são colocados da mesma forma que os intervalos de confiança, facilitando a interpretação. Este trabalho faz parte de uma coleção de vários livros, que contribuem para o entendimento das aplicações estatísticas.

TAGLIACARNE, G. **Pesquisa de Mercado: Técnica e prática**. São Paulo: Atlas, 1986, p.176/238.

Os conceitos de erro e nível de confiabilidade são tratados de forma simples neste trabalho. As seleções clássicas (probabilísticas) de amostragem e as não probabilísticas são colocadas de forma a instigar as investigações sócio-econômicas da área a ser amostrada, indicando a importância da seleção por quotas em pesquisa de opinião.

Para a formação de amostras, utilizando-se a seleção por áreas, o autor não exige áreas simétricas, isto é, a delimitação pode ser flexível, podendo adequar-se às várias necessidades práticas. Devido à falta de cadastramento de dados, a idéia de adequação de áreas é muito utilizada em pesquisas de opinião.

ADLER, Max K. **A Moderna Pesquisa de Mercado**. São Paulo: Pioneira, 1971.

Este trabalho contempla diversos aspectos sobre pesquisa de opinião, tais como o entendimento das variáveis em estudo, as formas de questionamento; amostragem probabilística e não probabilística.

Na amostragem por quotas, o autor compara os resultados da amostragem probabilística com os da não probabilística e afirma ser esta, em muitos casos, melhor que a amostragem probabilística, além de que os resultados são similares.

AGUIAR, Marco Antônio de Souza. **Manual Básico de Pesquisa de Mercado**. Brasília: Sebrae, 1998.

O autor, na realidade, faz uma síntese de como se fazer uma pesquisa de mercado, desde a identificação do problema, até se chegar às técnicas de pesquisa mais utilizadas. Os modelos de amostragem são apresentados de forma sintetizada, sem demonstrações. As ponderações nas aplicações feitas pelo autor são simples e de fácil entendimento.

NEUMAN, W. Russel. **The Paradox of Mass Politics, Knowledge and Opinion in the American Electorate**. Cambridge: Harvard University Press, 1986.

O autor analisa a sofisticação política sob três aspectos correlacionados entre si: a "saliência política", relacionada à preocupação e interesse pela política, que é mostrada por meio do interesse dos eleitores pelos noticiários; o grau de informação do eleitor sobre questões políticas e a capacidade de conceitualização da política, isto é, a capacidade de organizar idéias e "issues" por meio de conceitos ideológicos abstratos e de fazer uso do conhecimento político na avaliação dos "issues[1]". O nível baixo de informação do eleitorado, a falta de interesse, de conhecimento político e de capacidade de conceitualização dos fatos refletem o baixo grau de sofisticação política do eleitorado, que está associado diretamente ao nível sócio-econômico do eleitor. Apesar disso, o autor pondera e diz que nos países desenvolvidos, as políticas dão resultados positivos. Desta forma, fica clara a importância das variáveis sócio-econômicas quando se trabalha a opinião de pessoas em países subdesenvolvidos, onde os "ï ssues" não condizem com a realidade.

ELEIÇÕES 1998. [online] Disponível na Internet www.tre.gov.br.

Arquivo capturado em 07 de outubro 1998.

Este "site" do TRE fornece todos os resultados populacionais das eleições de 1998, por municípios e estado. As proporções a favor de cada candidato são fornecidas com relação ao comparecimento às urnas (veja anexo 04). Esse "site" também fornece os candidatos em ordem de votação por partido eleitoral.

IBIRITÉ [online] Disponível na Internet via www.almg.gov.br/munmg/M298.htm.

Arquivo capturado em 14 de Junho 1999.

Neste "site", as principais características sócio-econômicas do município de Ibirité são fornecidas. Não é um "site" abrangente, devido à falta de informações detalhadas sobre a distribuição populacional (faixa-etária, renda, nível de instrução e religião). Esta falta de informações, na realidade, gera dificuldades administrativas.

---

[1] Os issues são as demandas que se tornam objeto de maior atenção por parte das autoridades, em algum momento no interior do sistema político.

2. AMOSTRAGEM

2.1 BASE CONCEITUAL

A tentativa de estudar uma totalidade (população) poderia tornar-se inviável, devido a fatores relacionados ao tempo e aos custos. A solução encontrada para o estudo da população foi a retirada de uma amostra (parcela da população) que representasse a totalidade. Por meio do estudo da amostra, os dados são trabalhados e sintetizados com maior rapidez e os resultados podem ser utilizados para a resolução de problemas de necessidade imediata. Em determinados tipos de investigação, podem-se tornar necessários a utilização de uma equipe especializada e de equipamento tecnológico de primeira linha. A inspeção por amostragem surgiu nos laboratórios da Bell Telephone, determinando pesquisas pioneiras para aplicação da estatística. Shewhart (1910) fez um estudo sobre a inspeção de peças (milhares, de variados tipos, produzidas em série e montadas em conjunto no equipamento), no qual a inspeção completa era demorada e altamente onerosa, além de impraticável, pois os ensaios destruíam as peças. Apareceram, então, as técnicas de controle de processos e a redução da quantidade de peças inspecionadas. Nesse período, observou-se uma grande resistência dos engenheiros americanos, pois pensavam que as leis do acaso não tinham lugar adequado em meio a métodos científicos de produção. Com isto, o maior desenvolvimento das técnicas estatísticas se deu a partir da Segunda Guerra Mundial (Lourenço, 1980).

A amostragem pode levar a resultados mais exatos que a contagem integral da população, devido à redução do volume de trabalho (Cochran, 1977).

As principais fases de um levantamento por amostragem são:

a) Objetivos do levantamento;
b) População que fornecerá a amostra (população objetivo);
c) Dados (variáveis) que farão parte da amostra (grau de importância);
d) Grau de precisão desejado;
e) Métodos de medida;
f) Seleção das amostras;
g) Verificação preliminar;
h) Organização do campo de trabalho;
i) Sintetização e análise dos dados;
j) Informações básicas para futuros levantamentos.

Qualquer deficiência em uma das fases do levantamento por amostragem pode levar a resultados insatisfatórios, tendo como conseqüência soluções ineficazes.

Os objetivos principais da teoria de amostragem são: o aperfeiçoamento dos processos de seleção de amostras, os custos menores e as estimativas suficientemente precisas, visando os propósitos do estudo.

A precisão em um processo de amostragem é verificada através do exame da distribuição de freqüências, gerada pela estimativa, quando o processo é repetido, por diversas vezes, utilizando-se a mesma população (Cochran, 1977).

"Ao se fazerem inferências, o rigor não é alcançado com a eliminação completa da incerteza, mas pelo conhecimento explícito da mesma. A inferência estatística utiliza instrumentos para reduzir a incerteza e para lidar sabiamente com o que resta de certo" (Wallis & Roberts, 1964).

2.2 TIPOS DE AMOSTRAGEM

2.2.1 A Amostragem Aleatória Simples

A amostragem aleatória simples ou casual consiste em numerar a população e, por meio da escolha casual, selecionar os números dos elementos da população que farão parte da amostra. A escolha casual ou aleatória implica na mesma possibilidade de que cada membro da população possa ser o escolhido. A aleatoriedade é obtida por meio de algum processo mecânico ou eletrônico que tenha sido cuidadosamente verificado (veja anexo 01). Em muitos casos, em que se utilizam dados obtidos de fontes nas quais a dimensão espacial não é representada (cadastros, censos e outros), seleciona-se do total uma parte (amostra) de tal forma que o estudo se torne viável. A forma mais segura de se obter grupos similares é o processo aleatório ou casual. A aleatorização tende a separar os indivíduos segundo características similares, mesmo que o pesquisador não esteja consciente dessas características (Pião & Henry, 1995).

"No caso de unidades espaciais em que os pontos, trajetos ou áreas já estejam demarcados, é suficiente numerá-los de forma seqüencial e realizar a seleção seguindo a ordem de uma tábua de números aleatórios. Quando, porém, não há demarcação prévia das unidades da população, uma grade de

> coordenadas ortogonais numeradas a partir do canto sudoeste deve ser superposta ao mapa ou fotografia, para realização do sorteio. Selecionando-se de uma tábua de números aleatórios com quatro dígitos (os primeiros dois representam os valores de coordenadas leste e os dois últimos valores de coordenadas norte), e através destes localizamos os pontos que farão parte da amostra" (Gerardi & Silva, 1981).

2.2.1.1 Um exemplo de como se retirar uma a.a. simples da população.

Supondo que a população tenha 25 x 25 pontos, perfazendo um total de 625 elementos. Deseja-se retirar uma a.a. de 10 elementos.

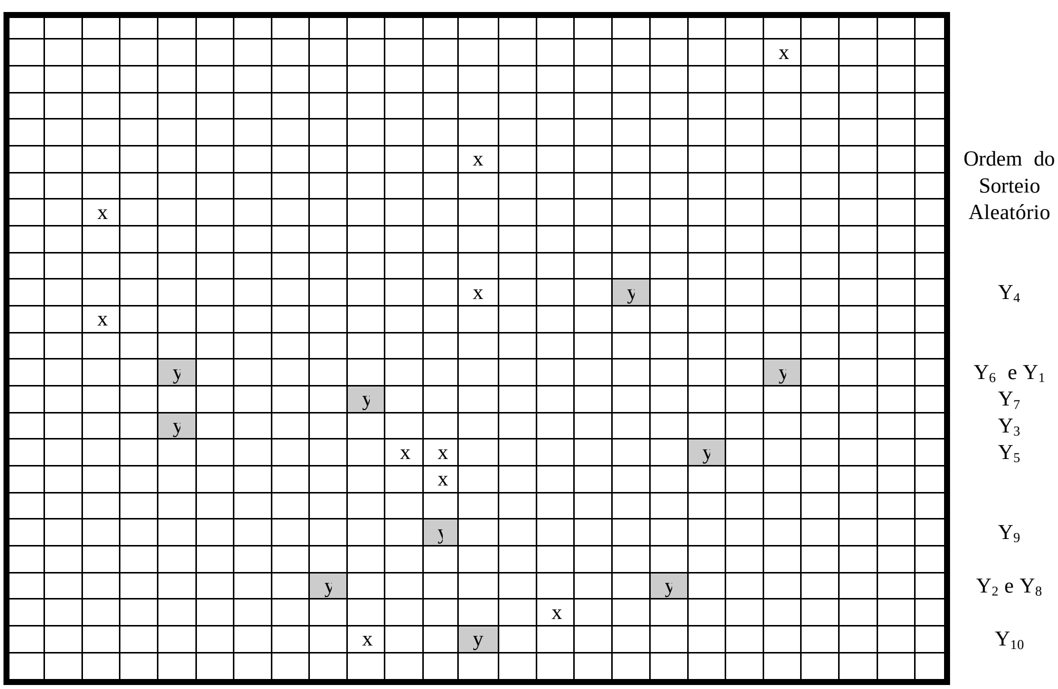
Primeiramente numera-se toda a população, caminhando da esquerda para a direita (na horizontal) e de cima para baixo. Na primeira fileira da tabela, os valores irão de 001 a 025, na segunda, de 026 a 050, ... e, na última, de 601 a 625. Vê-se a distribuição dos pontos na figura 01, em que cada célula representa um valor.

Depois que a população foi numerada, sorteia-se na TNA (Tabela de Números Aleatórios) o início de leitura (linha e coluna). Supondo que foi sorteado o valor 15 para a coluna e 26 para a linha, faz-se a leitura da TNA(L26,C15) no sentido vertical. Observa-se que para numerar a população foram necessários três dígitos, portanto, utilizaram-se três colunas para a leitura, a partir da coluna sorteada. Confere-se parte da TNA que se encontra no anexo 01.

26	722876	343520	597188	322037	458394	425390	102585
27	313882	180257	107158	392089	382208	628423	405873
28	538453	225472	619562	478439	729685	734884	307580
29	185906	725002	73**346**8	262954	615308	180607	653942
30	570995	225850	519650	150755	588017	188039	220153
31	527785	797163	35**524**8	865250	762726	062379	727674
32	373538	878412	29**380**7	611835	528422	203320	082313
33	374958	591277	68**267**7	159173	871361	153452	514538
34	872007	128734	636460	456386	468346	055975	345071
35	718657	129634	66**419**4	061052	319371	395292	850662
36	894767	159791	10**330**1	372063	799066	124560	398608
37	363284	196843	85**360**2	165334	887054	268779	706724
38	034998	491746	359540	166148	892098	743274	053289
39	738475	450657	41**542**2	349428	767497	798040	820590
40	723012	233648	47**487**0	095320	080681	708820	828704
41	270855	315395	65**588**5	497020	755538	364124	235176

Desta maneira, os elementos (y) sorteados para a amostra podem ser representados na seguinte forma:

FIGURA 03: Amostragem Simples



Fonte: EARLR / 1999.

Se a leitura fosse iniciada na TNA (L5;C2), a amostra seria diferente. Conferindo os pontos (x) na tabela anterior, fica claro que, apesar da diferença entre os pontos obtidos para as duas amostras, o mais importante é verificar se realmente a amostra representa bem a população. Outro fato interessante quanto a leitura na TNA é não haver repetição de elementos. Ao se fazer uma amostragem com ou sem reposição, os pontos serão os mesmos, quando a amostra for pequena em relação à população.

Outros métodos fornecem as mesmas probabilidades de um elemento ser sorteado como na amostragem simples, e alguns desses processos são mais úteis para certos problemas de amostragem. Os modelos de amostragem probabilística permitem levantar informações sobre conjuntos mais amplos de elementos - como universo ou população de referência - (Aguiar, 1998).

> "..., todos os planos de amostragem probabilística são estruturados a partir da amostragem aleatória simples. Antes de ser compreendido esse processo de "estruturação", devem ser compreendidas as técnicas de amostragem aleatória simples e as técnicas nela baseadas diretamente. ... . o âmago da questão da inferência estatística pode ser aprendido mais facilmente quando em relação a amostras aleatórias simples (Wallis & Roberts, 1964)".

2.2.2 A Amostragem Estratificada

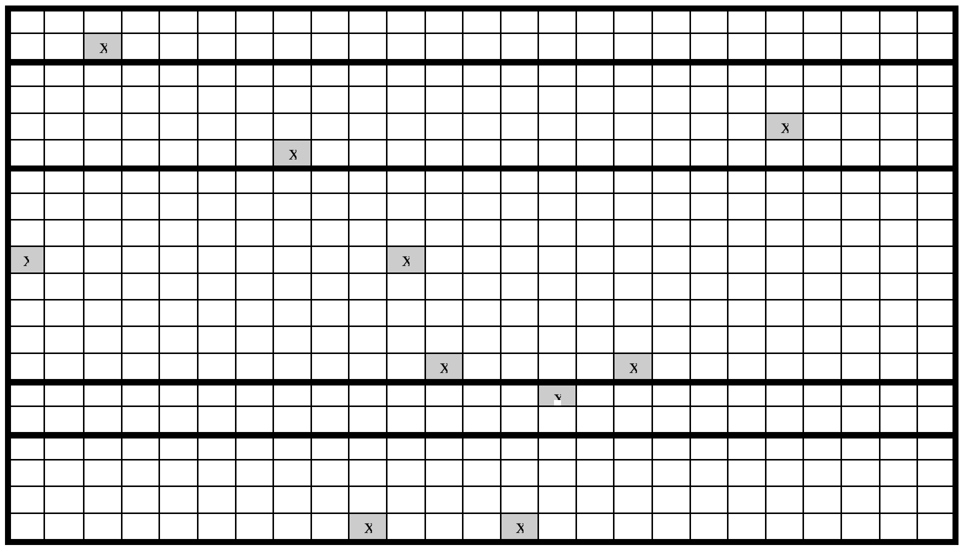
A amostragem estratificada é utilizada quando a população está distribuída em grupos heterogêneos entre si, isto é, grupos em que uma ou várias características relevantes são verificadas. Uma amostra aleatória simples é feita dentro de cada grupo, de forma que a proporção de cada estrato com relação à população total também esteja representada pela amostra.

> "... trabalhando com bacias hidrográficas podemos ter interesse em verificar características morfométricas em relação às ordens hortonianas e esse seria o critério de divisão dos estratos, ou ainda, na zona urbana, os estratos poderiam ser determinados por tipos de uso de solo ..." (Gerardi & Silva, 1981).

Para representar esse tipo de amostragem, suponha que a população tenha um total de 500 elementos divididos em cinco grupos heterogêneos entre si. As proporções de cada grupo em relação ao total sejam respectivamente de 10%, 20%, 40%, 10% e 20%, e que a amostra tenha de ter 10 elementos.

O primeiro passo é numerar internamente os elementos de cada estrato (os estratos estão separados por linhas pontilhadas (figura 04). Depois, escolhe-se através de uma amostragem simples, 1 elemento para o primeiro estrato, 2 elementos para o segundo estrato, 4 para o terceiro, 1 para o quarto e 2 para o quinto estrato. Os valores de início de leitura para os cinco estratos, respectivamente, são: TNA (L5; C10), TNA (L30;C40), TNA (L60,C27), TNA (L80;C3) e TNA (L73;C14). Assim, os elementos (x) para a amostra podem ser representados da seguinte forma:

FIGURA 04: Amostragem Estratificada



Fonte: EARLR / 1999.

2.2.3 A Amostragem à Base de Conglomerados

A amostragem por conglomerados será utilizada quando a população estiver distribuída em grupos homogêneos, isto é, grupos com as mesmas características. O procedimento consiste em escolher um ou mais dos conglomerados e retirar uma amostra aleatória simples. Existem também os casos de amostragem por conglomerados para mais de um estágio (Aguiar, 1998).

A amostragem de áreas é uma aplicação da amostragem à base de conglomerados, mas com outras características de planejamento. Para se obter uma amostra de probabilidade de populações amplamente dispersas, baseia-se na idéia de que os itens da população possam ser associados com áreas geográficas, extraindo-se uma amostra de probabilidades dessas áreas e fazendo-se a amostragem dentro delas. As partes da população nas áreas podem ser consideradas como conglomerados.

> "A região sobre a qual se estende a pesquisa é subdividida em pequenas áreas, uma para cada entrevistador. A delimitação dessas zonas é flexível e pode adequar-se a várias necessidades práticas. Por exemplo, uma das áreas poderá ser formada por um grupo de casas em um mesmo bloco de edifícios e constituir a zona de trabalho, por um dia, do entrevistador. Isto reduzirá o tempo e o custo da pesquisa." (Tagliacarne, 233, 1986).

2.2.4 A Amostragem Sistemática

A amostragem sistemática é utilizada quando a população encontra-se distribuída em ordem cronológica, alfabética ou temporal. O procedimento consiste em escolher um valor aleatório, $0 < b \leq r$, em que r represente a razão de amostragem (N/n = valor inteiro[2]), e escolher para a amostra os elementos cuja ordem são: b; b + r; ....; b + (n-1)r. Aqui, cada observação tem a mesma possibilidade de ser incluída na amostra, mas as probabilidades não são independentes.

Se a ordem dos elementos no cadastro for aleatória, a seleção sistemática será equivalente à amostragem aleatória simples. Mas se a ordem dos elementos do universo estiver correlacionada à variável principal de interesse, a seleção sistemática poderá fornecer melhores resultados que a aleatória simples, por introduzir um efeito de estratificação

---

[2] **N** representa o número de elementos da população, e **n** representa o número de elementos da amostra.

implícita. Se a ordem dos elementos da população for cíclica, a amostragem sistemática não deve ser utilizada, porque o ponto extraído para a amostra poderá coincidir com o período do ciclo ou ser múltiplo dele, introduzindo um grande viés na amostra (Aguiar, 1998).

Para exemplificar, pode-se supor que a população tenha um total de 200 elementos e que a amostra seja de dez unidades. Primeiramente, calcula-se a razão de amostragem $r = N/n = 20$. Utilizando-se a TNA(L88;C31), procura-se um valor $0 < b \leq 20$, então, $b = 18$. Portanto, as posições dos elementos (x) escolhidos para a amostra são: 018, 038, ... e 198, conforme estrutura abaixo.

FIGURA 05: Amostragem Sistemática

																	x							
												x												
							x																	
		x																				x		
																	x							
												x												
							x																	
		x																				x		

Fonte: EARLR / 1999.

2.2.5 A Amostragem Não Probabilística ou Não Casual

A amostragem não probabilística é aquela em que os elementos da população não têm a mesma chance de participarem da amostra, pois ela se baseia com exclusividade no que convém ao pesquisador. Um exemplo desse tipo de amostragem é a chamada amostragem por quota, na qual todas, ou a grande maioria das variáveis (características em estudo), são impostas como controle sobre a amostra. Apesar de não ser uma amostragem probabilística, oferece resultados similares ao de uma amostragem probabilística, sendo, em diversos sentidos, superior à mesma (Adler, 1971). Outro exemplo de amostra não casual é a seleção por julgamento ou conveniência nesta, o senso comum ou um julgamento equilibrado podem ser usados na seleção de uma amostra que seja representativa da população (Levin, 1978).

Em um trabalho sobre Geomorfologia Cárstica na região de Lagoa Santa, a unidade amostrada foi chamada de "Testemunho" e representava grande parte da evolução geomorfológica (milhares de anos). Logicamente, um único exemplar foi suficiente para o estudo (Kohler, 1989).

Em muitos casos, dependendo das características sobre as quais se desejam informações, principalmente características históricas evolutivas, a amostragem não probabilística será utilizada, e isso não quer dizer que a análise estatística será eliminada, pois as fases evolutivas poderão sugerir uma análise mais aprofundada dentro de um único exemplar.

Cochran (1997) exemplifica bem esse caso quando diz que para se fazer um exame de sangue, basta uma pequena amostra (supõe-se que o sangue esteja bem misturado), mas apesar de ser apenas uma unidade amostral, diversas características quanto à saúde podem ser analisadas a partir de uma única unidade amostral.

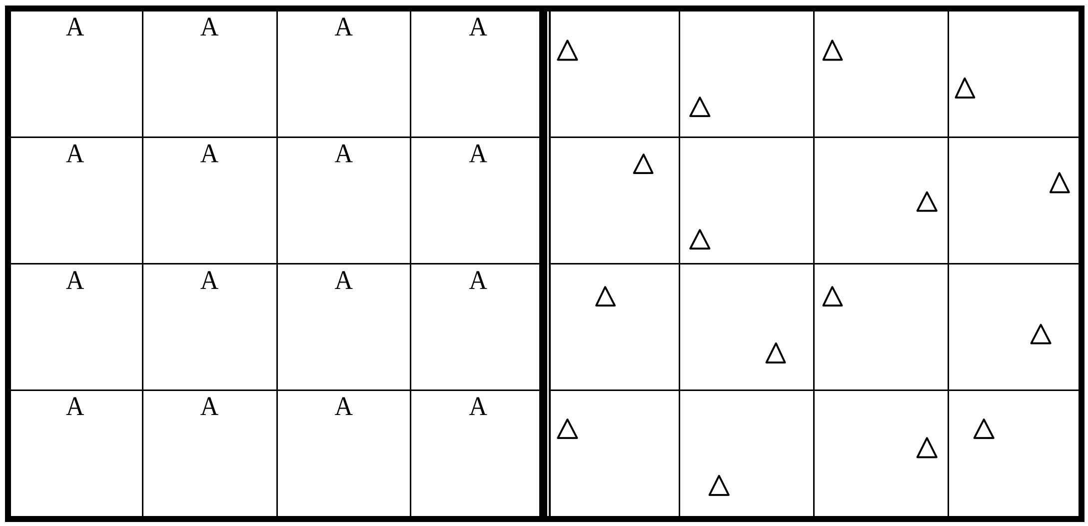
2.2.6 A Amostragem Estratificada Sistemática Não Alinhada

A amostragem estratificada sistemática não alinhada é utilizada quando se desconhece a distribuição geográfica dos fenômenos que se deseja estudar.

> "... desenvolvida por Berry e Baker (1968) corrige ou previne distorções decorrentes da amostragem sistemática alinhada e alia as propriedades da aleatoriedade e estratificação dando cobertura total à área amostrada ..." (Gerardi & Silva, 1981) .

Os elementos para uma amostragem sistemática não alinhada são encontrados da seguinte maneira: as coordenadas da unidade superior da esquerda são escolhidas em primeiro lugar, por meio de um par de números aleatórios. Outros três números aleatórios determinam as coordenadas horizontais das unidades restantes da primeira coluna de estratos. Outros três números aleatórios são necessários para a fixação das coordenadas verticais restantes da primeira fileira dos estratos. O intervalo constante K (igual ao lado dos quadrados) determina, conseqüentemente, a localização de todos os outros pontos. As pesquisas de Quenouille (1949) e de Das (1950) sobre correlogramas simples bidimensionais indicam que o sistema desalinhado é freqüentemente superior tanto à grade quadrada, quanto à amostragem aleatória simples. A primeira parte da figura 06 representa uma grade quadrada em que a amostra fica inteiramente determinada pela escolha de apenas dois dígitos aleatórios (coordenadas da unidade superior da esquerda). A segunda parte da figura 06 representa uma amostra aleatória desalinhada. Para esse tipo de amostragem, supõe-se que cada quadrícula está dividida em cem (100) pequenas quadrículas, sendo necessário o uso do sistema cartesiano para posicionar os pontos. Existe, também, a amostragem sistemática centralizada em que os pontos para a amostra serão representados pelos pontos centrais de cada estrato.

FIGURA 06: Amostragem Sistemática Alinhada e Não Alinhada



Fonte: EARLR / 1999.

2.2.7 A Amostragem Painel

Chama-se painel a uma mesma amostra observada continuamente em intervalos de tempo regulares e com pelo menos um tema principal repetido. Este tipo de amostra permite verificar modificações de comportamento sob imposição de novas estruturas. Devido às constantes modificações nas variáveis demográficas e sociais básicas, faz-se uma rotatividade do painel, que será definida, observando-se as modificações na estrutura das variáveis principais. Esse tipo de amostra é muito utilizado em pesquisas sociais, políticas e de mercado (Aguiar, 1998). Outro exemplo de amostragem painel refere-se ao estudo dos lençóis freáticos, onde diversos postos são monitorados frequentemente, para se verificar a quantidade de água disponível.

Apesar da vasta utilização da amostragem aleatória, dependendo do estudo, existe a necessidade de se combinar diversos tipos de seleção, para se obter uma amostra que realmente capte as principais características populacionais.

Com vistas à inferência estatística, uma amostra probabilística é o procedimento de amostragem exigido. Embora uma amostra probabilística possa conduzir à uma seleção não representativa da população, é sempre possível especificar a precisão provável dos resultados.

A amostragem aleatória simples constitui o método mais comumente usado na coleta de dados, embora outras variações, às vezes, se mostrem necessárias e desejáveis na prática. (Burt & Barber, 1996).

As estatísticas de uma amostra são variáveis aleatórias. O valor de qualquer estatística de amostra varia de amostra para amostra para todas as amostras de uma dimensão fixa (n) retiradas de uma mesma população. Para amostras suficientemente grandes, a distribuição das estatísticas amostrais é aproximadamente normal (veja anexo 02).

Por ser de importância relevante, nesse estudo, a amostragem espacial será trabalhada no próximo tópico.

3. AMOSTRAGEM ESPACIAL

3.1 FENÔMENO ESPACIALMENTE DISTRIBUÍDO

Quando para se fazer uma pesquisa ou um estudo de caso, existe a necessidade de coleta de dados, o planejamento não é de simples execução , devido, em muitos casos, à falta de cadastramento e espacialização (localização) das variáveis de interesse.

Uma seleção de amostragem deve acompanhar a distribuição da população. Um estudo de caso em campo exige um estudo das seleções de amostragem e da inferência estatística para orientar os trabalhos de espacialização das variáveis (veja anexo 02).

A espacialização das informações, tanto para a coleta de dados quanto para a análise de fenômenos, é uma ferramenta que, cada vez mais, torna-se necessária para o desenvolvimento das várias áreas.

Na maioria dos casos, quando se faz uma seleção de amostragem, o espaço é trabalhado como se fosse um fenômeno secundário, e não se parte do espaço para criar a seleção de amostragem. Para se ter uma visão mais direcionada das seleções de amostragem no espaço, é necessário verificar as seleções de amostragem geográficas.

3.2 AMOSTRAGEM ESPACIAL

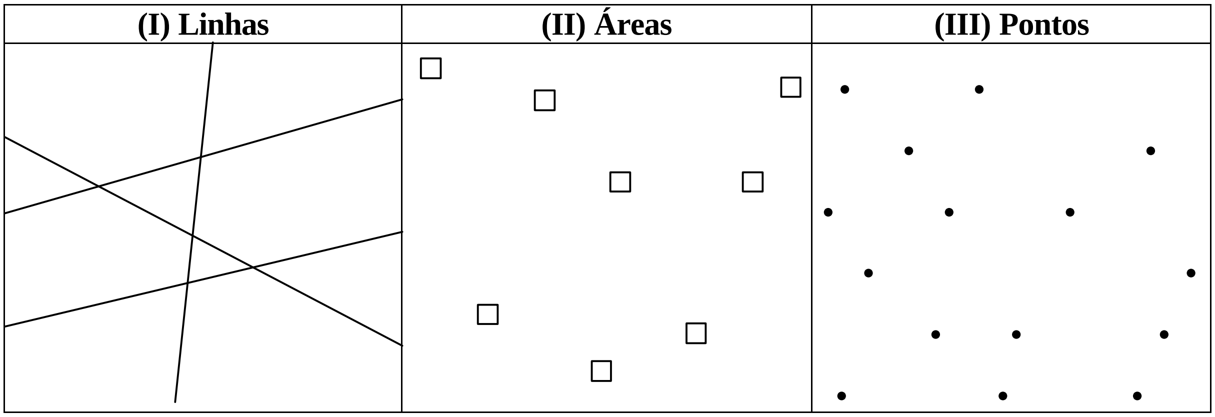
"Para o geógrafo, há pelo menos dois motivos para que se faça uso de amostras espacialmente distribuídas. Primeiro, há o caso em que o geógrafo deve ir a campo e amostrar algum fenômeno espacialmente distribuído. Um biogeógrafo pode desejar estudar a distribuição de uma espécie de planta ou mesmo a distribuição de sementes de uma planta ou conjunto de plantas. A amostragem em campo também pode ser necessária para a confecção de um mapa de solo. A amostragem de pontos localizados em um mapa, onde o fenômeno esteja continuamente distribuído, é muito importante. Mapas de uso terrestre das cidades, mapas de solo e mapas de vegetação são exemplos típicos desta forma de amostragem.

Para os fenômenos distribuídos discretamente num mapa, podem-se usar os procedimentos clássicos de amostragens. Para os padrões complexos existentes na maioria dos mapas, cujas variáveis são contínuas, tal procedimento é completamente inapropriado. Ao invés disso, esses mapas são amostrados, usando-se amostras de linhas, áreas ou pontos (veja a figura 07).

Na amostragem de um mapa, basta utilizar o sistema de coordenadas cartesianas para identificar qualquer ponto e, conseqüentemente qualquer elemento da população. Este sistema de coordenadas abrange todas as necessidades de uma estrutura de amostragem, ou seja, para cada elemento da população existe apenas um endereço (uma localização). Os pares de coordenadas (x , y) são mutuamente excludentes, pois não existem dois ou mais pontos do mapa com coordenadas idênticas. Os cartógrafos reconhecem que problemas de distorção espacial, associados a mapas de pequena escala, são ignorados.

Considerando o sistema de coordenadas (figura 08), usado para se obter amostra de áreas. Os eixos seguem as orientações usuais norte-sul e leste-oeste, com origem na parte inferior esquerda ou sudoeste. Os eixos estão subdivididos em unidades para qualquer grau de precisão desejado.

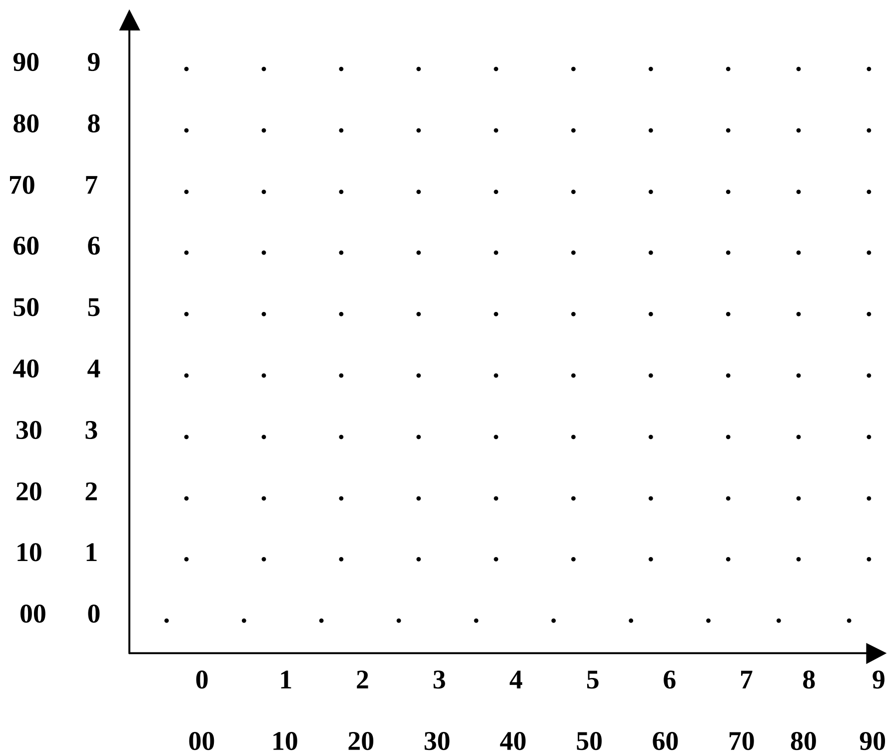
FIGURA 07: Amostragem por Linhas, Áreas e Pontos



Duas possibilidades estão apresentadas na figura 08. A grade de um dígito usa os números de 0 a 9 para identificar pontos particulares, tanto no eixo x, quanto no eixo y. Desta forma, os números aleatórios podem ser usados para localizar qualquer ponto no mapa. O primeiro número, escolhido aleatoriamente, é usado para definir a coordenada x e o segundo para a coordenada y. Se apenas números aleatórios de um dígito forem usados, somente os cem (100) pontos mostrados, como em seções de grade, podem realmente ser amostrados. A população disponível é bem diferente da população alvo ou total, que constitui todo o mapa.

Se números aleatórios de dois dígitos forem utilizados para a grade em cada eixo, haverá 10.000 pontos possíveis para serem incluídos na amostra. A grade usada para definir a estrutura de amostragem deve ser suficientemente exata, de tal forma que as características de apenas algumas áreas pequenas sejam representativas da população disponível. Dependendo do mapa, pode ser necessário o uso de referências de grade de 2, 3 ou mesmo 4 dígitos, a fim de que o conjunto de pontos disponíveis cubra adequadamente o mapa alvo " (Burt & Barber, 1996)[3].

FIGURA 08: Sistema de Coordenadas para Amostragem de Áreas



3.2.1 Amostragem de Áreas e Linhas

"Um dos métodos de amostragem normalmente mais usados é a amostragem de áreas. Esta se difere significativamente do sistema utilizado no exemplo anterior. Na figura 09, três padrões diferentes de projetos de amostragem estão ilustrados:

a) Considerando o projeto aleatório de amostragem de área em (I). Tal projeto pode ser desenvolvido, utilizando-se três regras simples: • seleciona-se uma área; • escolhe-se o

---

[3] Tradução Livre.

tamanho da área (quanto maior a área mais difícil é a análise) e; • se for necessário, calcula-se a proporção da área em diversas categorias de uso de terra, que se torna uma tarefa longa, se a área for grande.

b) Selecionam-se dois números aleatórios, em que o primeiro localize a coordenada x da área objeto de estudo e o segundo localize a coordenada y. O ponto (x,y) representa o centro da área.
c) Selecionam-se quantas áreas forem necessárias para alcançar o tamanho estimado suficiente para a amostra.

Como variação deste algoritmo, é possível orientar aleatoriamente as áreas, selecionando-se um terceiro número aleatório para controlar tal orientação, mostrada na figura 09 (II). Para muitos mapas complexos, a distribuição irregular do padrão de mapa dentro de uma área pode tornar a análise muito difícil. Em tal caso, a amostragem do local pode ser uma alternativa desejável.

Para se fazer uma rotação de áreas, pode-se utilizar o sistema de graus. Se a área for simétrica, escolhem-se dígitos aleatórios[4] de 000 a 180. Se a área for assimétrica, escolhem-se dígitos de 000 a 360. Estes dígitos representarão o número de graus a serem rotacionados (veja figura 10). O sentido escolhido foi o sentido horário, pode-se, também, ser utilizado o sentido anti-horário[5].

---

[4] Devido à simetria, as áreas seriam as mesmas a partir de 180^o, isto é, a rotação de 1^0 corresponde à mesma área de uma rotação de 181°

FIGURA 09: Amostragem de Áreas e Linhas

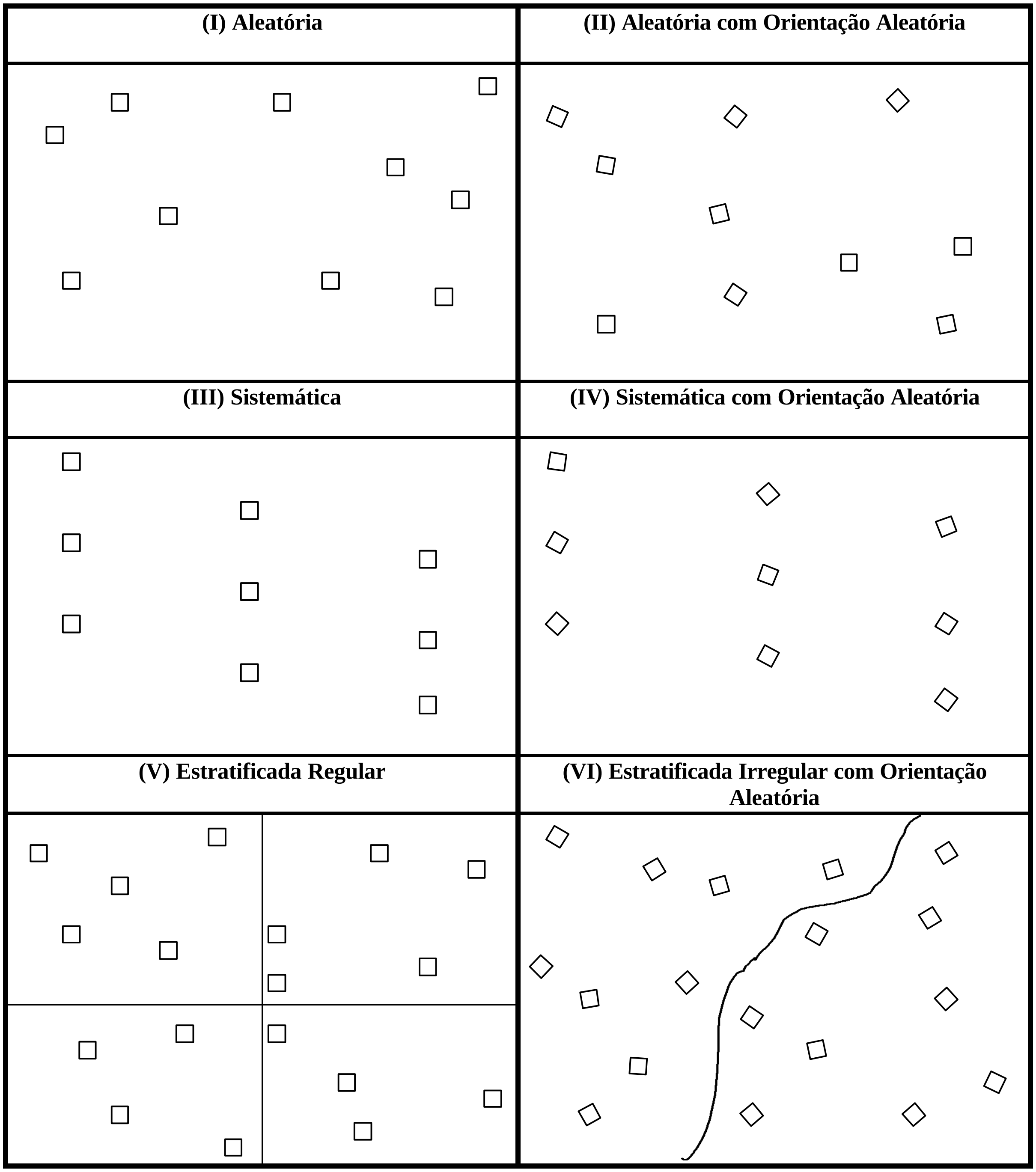
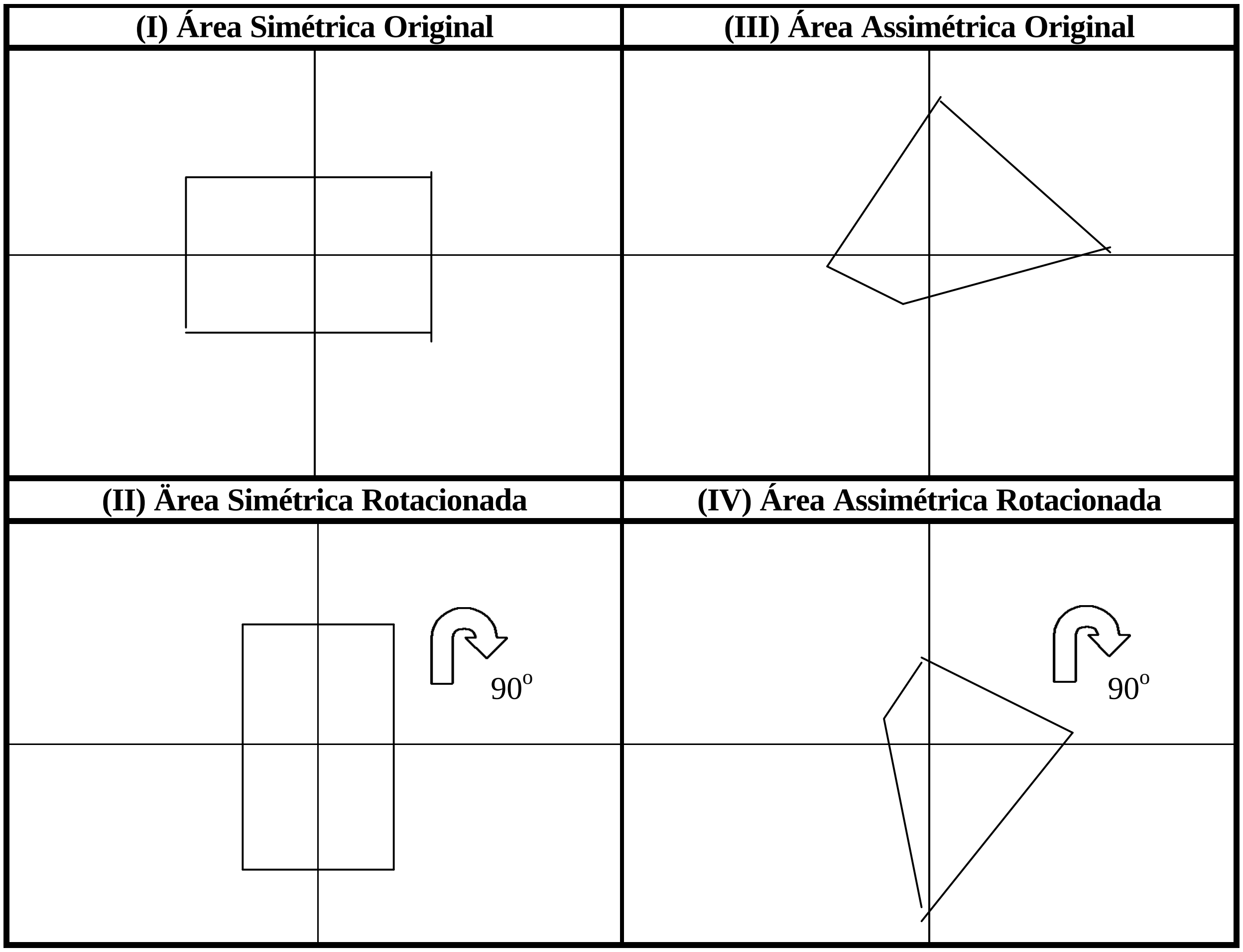


Figura 14: Os quadrados(□) representam as unidades amostradas.

---

[5] O sistema de graus pode ser associado ao sistema de coordenadas geográficaa (norte/sul e leste/oeste).

FIGURA 10: Rotação de Áreas Simétricas e Assimétricas.



Para uma amostragem de linhas, supõe-se uma amostra de L linhas de extensão total u, para um projeto de amostragem aleatória, realizado do seguinte modo: • para especificar um ponto no limite do mapa que possa ser utilizado como ponta da primeira linha, escolhe-se um número aleatório de dois dígitos simples; • escolhe-se um segundo número de dois dígitos para localizar um ponto no limite para a outra ponta da linha; • ligam-se os dois pontos e determina-se a primeira linha; • determina-se a extensão dessa linha.

Para completar a amostra é necessário amostrar uma extensão de $(L - l_1)$u, isto é, a diferença da extensão total menos a extensão da 1ª linha. Assim, mais dois números aleatórios são usados para localizar uma 2ª linha e a extensão da amostra de linha é atualizada para $(L - l_1 - l_2)$u. Este processo é repetido até que Lu de linha tenha sido selecionado para a seleção de amostragem"[6].

Para exemplificar, suponha que se deseje estimar o percentual de árvores em uma floresta de 100 km^2, utilizando a amostragem por linhas, e que serão amostrados 10 km de linhas. Primeiramente, deve-se colocar o sistema de coordenadas sobre o mapa e escolher o ponto de partida.

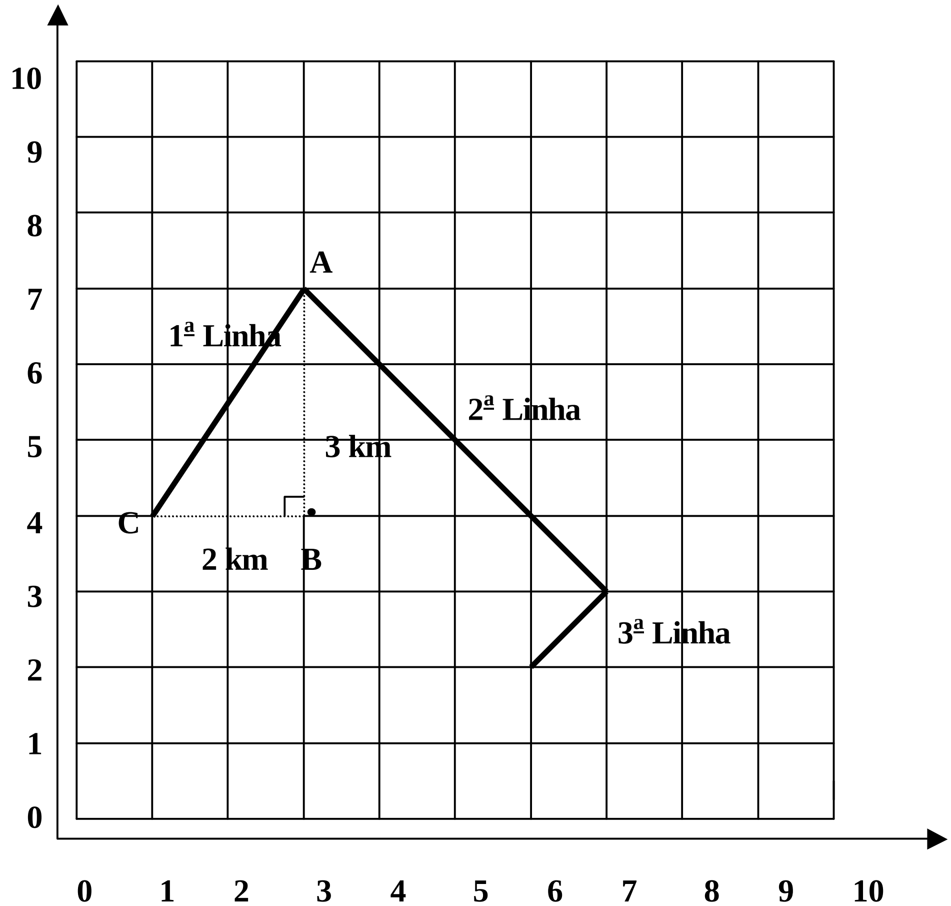
A seleção será aleatória simples, pois os outros casos são extensões da mesma. Agora deve-se escolher dois dígitos aleatórios para a o início da 1ª linha, e mais dois dígitos aleatórios para o final da linha (que será o início da 2ª linha). Suponha que os valores escolhidos aleatoriamente para o início da 1ª linha foram (1 ; 4), e para o final da linha (3;7). A figura 11 representa a primeira linha e sua extensão, que pode ser calculada, utilizando-se o Teorema de Pitágoras (veja figura 12).

Caso o total encontrado nas três linhas fosse de 5.533 árvores, a proporção destas na floresta seria de 553,3 (árvores por km) ou 0,55 árvores/m.

Outra maneira de se fazer amostragem por linhas é a escolha de cada linha independentemente uma da outra, para isto basta escolher dois pares de números aleatórios para cada linha, um par para o início e outro para o término da linha.

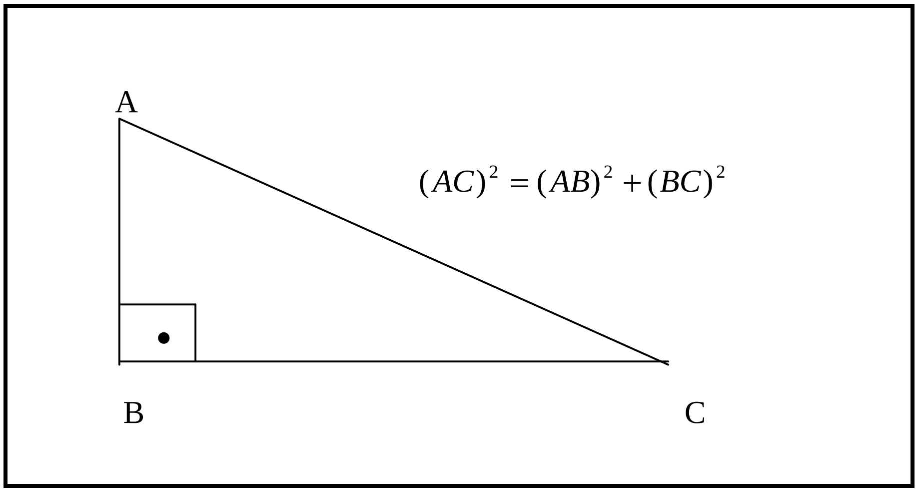
[6] Tradução Livre.

FIGURA 11: Amostragem Por Linhas.



Desta forma, cada quadrícula terá 1 km^2, e a primeira linha terá uma extensão de $\sqrt{3^2 + 2^2} \cong 3{,}61$ km. Escolhendo-se outro par de números aleatórios, por exemplo, os valores (7 ; 3), teremos a segunda linha cujo comprimento é de aproximadamente 5,66 km. Para um comprimento total de 10 km, ainda faltam (10 - 3,61 - 5,66) = 0,73 km a serem amostrados. Assim, escolhe-se outro par de valores aleatórios (6 ; 2) que fornecerá uma linha de 1,41 km. Observa-se que a amostra ultrapassou os 10 km, assim, ou se amostra toda a linha, ou 0,73 km da terceira linha.

FIGURA 12: Teorema de Pitágoras.



"O projeto de amostragem sistemática da figura 13 (II) é baseado em um padrão regular de linhas perpendiculares, com um intervalo preestabelecido. Supondo que uma amostra de extensão L = 250u de linhas deva ser retirada de uma área de $25u^2$, haverá cinco (5) linhas em cada uma das direções norte-sul e leste-oeste cada uma com 25u de comprimento. Usa-se um intervalo de 5u entre as linhas. Assim, um número aleatório simples pode ser usado para localizar as linhas leste-oeste e um outro para localizar as linhas norte-sul. Se houver periodicidades no mapa, isto é, características lineares ou quase lineares, o procedimento de amostragem sistemática pode levar a estimativas tendenciosas de algumas características do mapa.

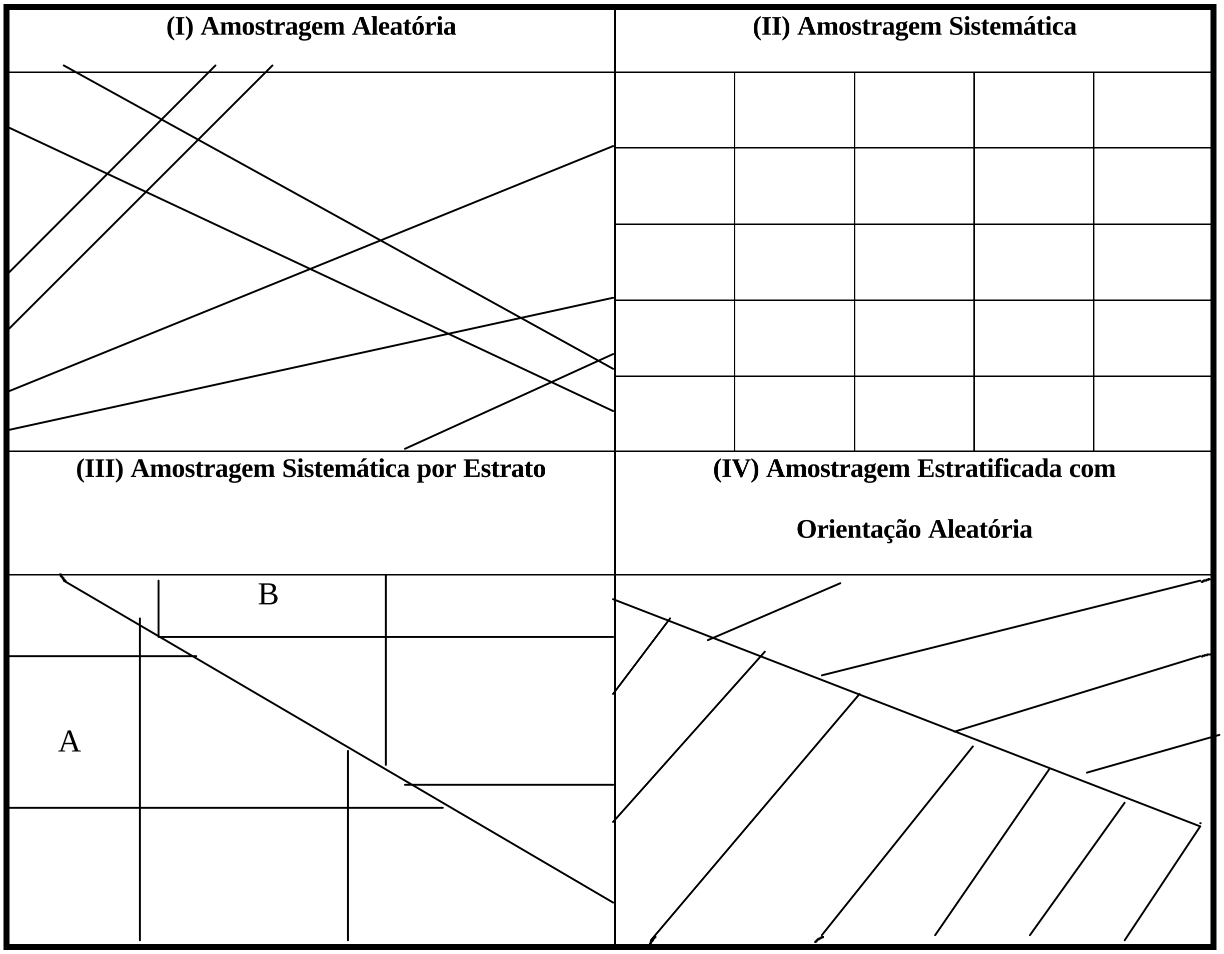
Ruas, auto-estradas, redes pluviais e outras características similares de mapas podem levar a erros significativos de amostragem quando os mapas são amostrados usando-se um projeto sistemático.

É possível orientar aleatoriamente as linhas de uma amostra sistemática, embora isto seja de difícil operacionalização. Modificações adicionais são também necessárias para amostras de uma área de forma irregular. Entretanto, significativo ganho de tempo pode ser obtido da amostragem sistemática. Frequentemente é bastante válida a complexidade acrescentada.

Um projeto de amostragem estratificada de área para um mapa irregular encontra-se na figura 13 (III). As áreas ou estratos podem ser representados por curvas de nível definidas, características geomorfológicas ou geológicas, ou até mesmo por limites municipais de uma área urbana. Dois estratos,A e B, são retirados da figura 13 (III) na proporção 60:40. Desta forma, 60% da extensão da linha devem ser retirados da área A, e 40% da área B. Inicialmente, números aleatórios são escolhidos para localizar as linhas individualmente. Uma vez que a extensão das linhas do estrato A alcança 60% da extensão da linha da amostra desejada, somente as partes das linhas subseqüentes do estrato B são usadas para completar a amostra.

Se as linhas, inicialmente, preencherem a extensão desejada do estrato B, as extensões das linhas subseqüentes de B serão ignoradas e somente as extensões de A serão incluídas na amostra. Este procedimento pode ser estendido para se manipular qualquer número de áreas ou estratos, de forma regular ou irregular. Para formas regulares, pode ser mais fácil amostrar cada estrato individualmente. Linhas lineares para mapas estratificados por áreas podem também ser tomadas de uma maneira sistemática e com uma orientação aleatoriamente escolhida como mostra a figura 13 (IV)" (Burt & Barber, 1996)[7].

FIGURA 13: Projetos Para Amostragem de Linhas



Fonte : EARLR / 1999.

3.2.2 Amostragem por Pontos

"Talvez a forma mais simples de amostragem espacial seja a amostragem por pontos. Selecionar uma a.a. simples de pontos de um mapa não é realmente mais difícil do que selecionar uma a.a. de uma estrutura clássica. Para se identificar cada ponto de uma amostra, são necessários dois números aleatórios: um é usado para a coordenada x, e outro

---

[7] Tradução Livre.

para a coordenada y. Dois números aleatórios são escolhidos para cada ponto de uma amostra, até que "n" pontos tenham sido localizados. A figura 14 (I) mostra um padrão aleatório típico de pontos. Normalmente, existem algumas áreas do mapa em que nenhum ponto é amostrado, e algumas áreas cuja densidade de pontos é elevada. Uma vez que o projeto é meramente aleatório, há sempre a possibilidade de que se possa retirar uma amostra não representativa. Isto se deve à suposição básica de amostragem clássica simples de que todos os elementos da população têm as mesmas características.

Para se assegurar de que todas as partes do mapa foram amostradas, e de que não há nenhuma alta concentração de pontos aleatórios, pode-se retirar uma amostra aleatória sistemática de pontos (veja figura 14 (II)). Os dezesseis pontos (16) para esta amostra particular são regularmente espaçados em intervalos eqüidistantes em termos de coordenadas x e y. Este projeto é de fácil operacionalização. O intervalo é escolhido de tal forma que o número de pontos escolhidos no mapa se iguale ao tamanho da amostra. Para se fazer este tipo de seleção, primeiramente, escolhe-se aleatoriamente o ponto do canto sudoeste do mapa, dentro da pequena área delimitada pelas linhas interrompidas (quadrícula). Então outros pontos são escolhidos em outros intervalos fixos.

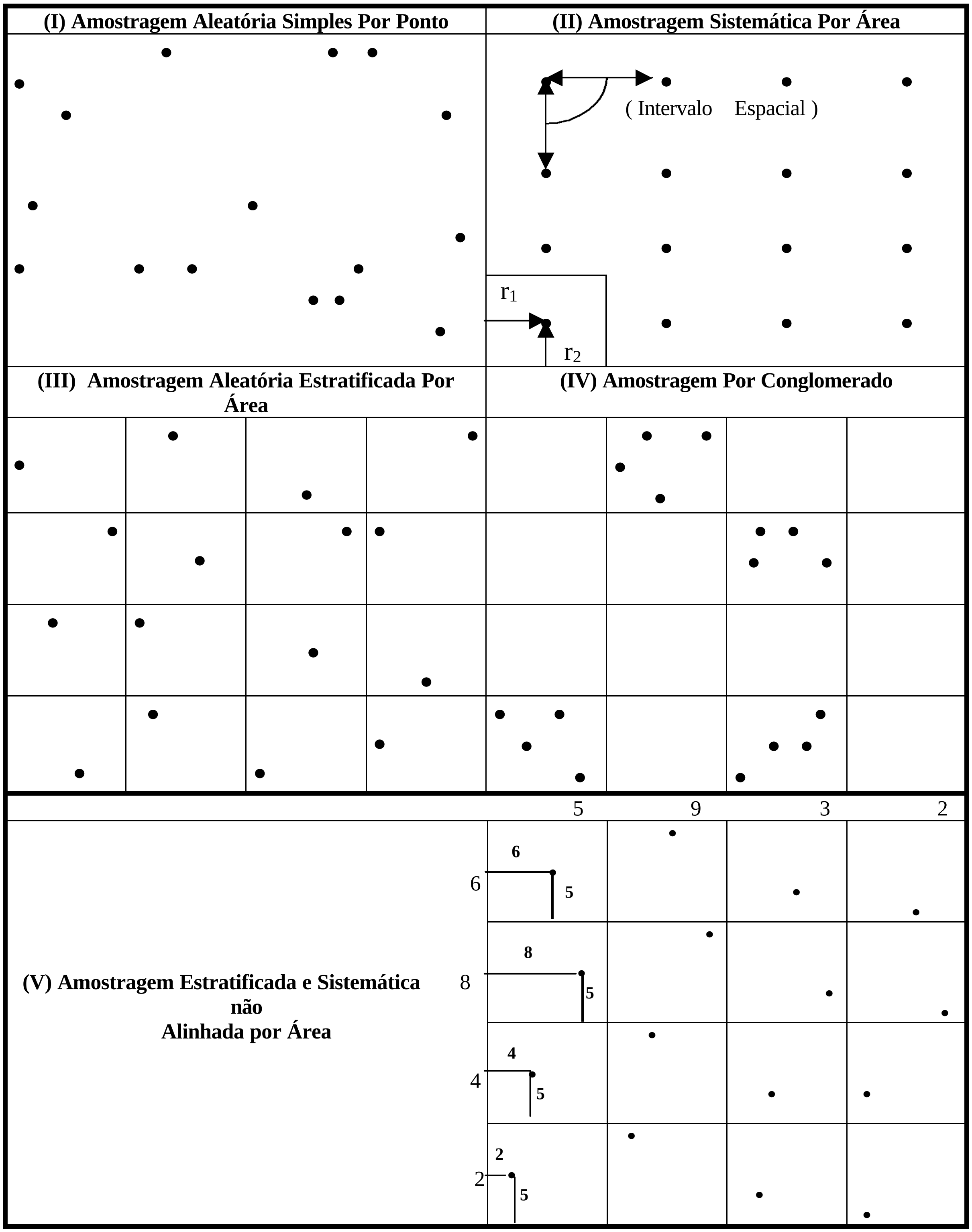
Somente dois (2) números aleatórios são necessários para se fazer uma amostragem sistemática de pontos e isto independe do tamanho da amostra. Uma vez que a localização de um ponto simples do projeto seja conhecida, a localização de todos os outros pontos é igualmente conhecida. Como este projeto seleciona pontos em intervalos regulares, o projeto sistemático pode levar a uma estimativa tendenciosa da proporção de um mapa coberto por algum fenômeno. A amostragem sistemática é amplamente utilizada em estudos de solo e vegetação.

Quando se usa a amostragem aleatória estratificada espacial, ela tem a mesma vantagem da amostragem sistemática, isto é, cobre quase todo o mapa (como ilustrado na figura 14 (III)).

Os projetos estratificados têm menos erros de amostragem do que os projetos aleatórios simples ou sistemáticos. Uma amostra por conglomerados pode também ser extraída de um mapa que tenha sido estratificado por áreas. Como se observa na figura 14 (IV), o procedimento usual é selecionar aleatoriamente certas áreas do mapa, e então amostrar aleatoriamente dentro dos conglomerados. A vantagem dessa amostragem, especialmente na amostragem de campo, deve ser avaliada contra a possível desvantagem de se omitirem grandes áreas do mapa.

Um projeto combinado, que incorpora elementos destes projetos mais simples, é a amostragem estratificada sistemática não alinhada da figura 14 (V). Ela é estratificada, desde que o mapa seja dividido em células (quadrículas) antes da amostragem e sistemática, porque é desnecessário usar dois números aleatórios para a escolha de cada ponto da amostra. Entretanto, ao contrário do caso de uma amostra sistemática simples, o padrão advindo deste procedimento não é alinhado. Para desenvolver este raciocínio, define-se uma grade simples (x,y) dentro de cada quadrícula. Um número aleatório simples é escolhido para cada linha, e outro para cada coluna. O número aleatório da linha define a coordenada x dentro de cada quadrícula dessa linha, e o número aleatório da coluna localiza a coordenada y para todas as quadrículas da coluna. As localizações de alguns pontos estão mostradas na figura 14 (V).

FIGURA 14: Amostragem Por Pontos



Fonte: EARLR / 1999.

A escolha do projeto na amostragem espacial depende diretamente do fenômeno a ser estudado. Cada projeto espacial compartilha as características básicas de sua contraparte não espacial. Se o mapa escolhido tiver características sistemáticas, um projeto sistemático representará uma escolha indevida. Se não houver um padrão óbvio no mapa, todos os projetos fornecerão resultados bem similares. Tipicamente, as amostras estratificadas são mais precisas que as amostras aleatórias simples, que, por sua vez, são mais precisas que as amostras por conglomerados. Uma pesquisa experimental para testar a precisão de todos os projetos de amostragem, mostrou que o projeto aleatório estratificado sistemático não alinhado é decididamente superior a todos os outros projetos" (Burt & Barber, 1996)[8].

Em síntese, as amostras podem ser selecionadas a partir de uma população de muitos modos. Normalmente, elas são retiradas de modo a minimizar tanto o tempo quanto o custo dos dados na coleta. Geralmente, quatro passos precedem a coleta real de dados:

- a especificação da população;
- o delineamento da estrutura de amostragem;
- a escolha de um projeto de amostragem e;
- um pré-teste dos procedimentos de coleta de dados.

Os três tipos de amostragem (pontos, linhas e áreas) sob um determinado espaço geográfico estão ilustrados por meio das figuras 15, 16 e 17. O espaço a ser amostrado é representado por dois bairros da cidade de Ibirité e será denotado por R2A.

---

[8] Tradução Livre.

FIGURA 15: Amostragem Por Pontos - R2A - Cidade de Ibirité.

Fonte: EARLR / 2000

Observe que dos dez pontos amostrados, alguns estão fora da área urbanizada. Assim, se o estudo não comportar estas áreas, deve-se amostrar novos pontos. Se existir interesse nos dois tipos de áreas (urbanizadas e não urbanizadas), a seleção pode ser por pontos, mas estratificada.

FIGURA 16: Amostragem Por Área - R2A - Cidade de Ibirité.

Fonte: EARLR / 2000

O mesmo procedimento da escolha aleatória simples deve ser feito. Outro fato importante é que, em áreas urbanas, a simetria não é perfeita, principalmente em cidades não projetadas. Assim, na prática, existe a necessidade de um ajuste, sendo usual, a numeração das áreas para uma escolha aleatória.

FIGURA 17: Amostragem Por Linhas - R2A - Cidade de Ibirité.

Fonte: EARLR / 2000

Nos três planos amostrais fica clara a importância do conhecimento das características da população. Apesar da vasta utilização da amostragem aleatória, dependendo do estudo, existe a necessidade de se combinar diversos tipos de seleção, para se obter uma amostra que realmente capte as principais características populacionais.

Com vistas à inferência estatística, uma amostra probabilística é o procedimento de amostragem exigido. Embora uma amostra probabilística possa conduzir à uma seleção não representativa da população, é sempre possível especificar a precisão provável dos resultados.

A amostragem aleatória simples constitui o método mais comumente usado na coleta de dados, embora outras variações, às vezes, se mostrem necessárias e desejáveis na prática (Burt & Barber, 1996).

Como enriquecimento do que foi tratado até aqui, o próximo tópico mostra um projeto de seleção de amostragem em que os dados iniciais não estão georeferenciados, mas o tratamento proposto para a seleção de amostragem, apesar de não ser probabilístico, oferece excelentes resultados.

4. ESTUDO DE CASO

4.1 AMOSTRAGEM POR QUOTAS - CIDADE DE IBIRITÉ

Ibirité é um município mineiro de pequeno porte, situado na Região Metalúrgica, a uma distância de 21 km de Belo Horizonte, por rodovia. A área do município é de 145 km^2, com o relevo distribuído da seguinte forma: 10% plano, 60% ondulado e 30% montanhoso. É servido pelo Ribeirão Ibirité e Córrego Capão da Serra.

Em 1996, o total de arrecadação do município foi de três milhões duzentos e oitenta e nove mil e duzentos e três reais (ICMS e outros impostos). Da população de mais de cem mil habitantes, cerca de 15% está ocupada e distribuída entre vários setores econômicos: Agropecuário, Industrial, Comércio de Mercadorias, Transporte, Comunicação e Armazenagem, e Outros Serviços.

Os principais produtos agrícolas da cidade de Ibirité são: arroz, cana de açúcar, feijão, mandioca, milho, tomate e cebola. A pecuária se distribui entre a criação de bovinos, eqüinos, suínos e galináceos.

O principal consumo de energia elétrica da cidade é o residencial, 30% maior que a soma dos consumos industrial, comercial, rural e outros. A sede do município possui apenas um hospital e dois centros de saúde. Os principais órgãos públicos são dependentes de Belo Horizonte (IBIRITË; 1999). Ibirité é um município pobre, porém as condições econômicas da população não interferirão diretamente na aplicação da seleção das amostras, visto que as variáveis de controle foram preestabelecidas, procurando controlar esse fator. A escolha da cidade de Ibirité, para o estudo de caso, está relacionada com a distância (acesso) de Belo Horizonte, (veja figura 18).

O propósito da pesquisa seria o de verificar se houve diferença significativa nas tendências quanto à preferência dos eleitores, utilizando-se os resultados das três últimas seleções de amostragem e o resultado oficial do TRE. Para se verificar as intenções de votos para a última eleição, foram feitas diversas seleções de amostragem em Ibirité. As três últimas amostras têm os mesmos procedimentos, mas as duas últimas têm uma estratificação a mais. Durante esse estudo, pode-se verificar que todo o desenvolvimento teórico está diretamente relacionado às amostragens clássicas.

A amostragem por quotas foi escolhida por ser um tipo de amostragem não probabilística, tendo em vista ser desconhecida a distribuição populacional real.

Os dados disponíveis para esse estudo sobre a opinião dos eleitores de Ibirité foram disponibilizados pelo IBGE, TRE e pela Pólis Consultoria. Com base nesses dados foi feita a seleção das amostras. Os dados do TRE estão colocados sob a forma de Região Eleitoral (agrupamento de bairros). Os dados do IBGE estão relacionados com as características gerais da população (sexo, faixa etária e nível de instrução). Observa-se que há apenas a cidade e seus bairros, no traçado urbano da cidade (anexo 06), que poderiam assumir a forma georeferenciada, mas para cada eleitor isto não é possível. Os dados do TRE e do IBGE estão no anexo 05.

4.2 SELEÇÃO DAS AMOSTRAS

Como a primeira fase em uma seleção de amostragem numa pesquisa de opinião (como em qualquer pesquisa em campo) é a determinação da localização da população a ser amostrada, neste caso, ela é representada pelo traçado urbano da cidade.

Pelo TRE, a população da cidade de Ibirité, que será amostrada, está dividida em oito (8) regiões eleitorais, conforme tabela 01.

TABELA 01: Eleitorado de Ibirité por Região - 1998

Regiões	Nº de eleitores	%
1	9.848	19.4
2 A	3.096	6.1
2 B	5.533	10.9
3	2.893	5.7
4	1.421	2.8
5	14.416	28.4
6	3.452	6.8
7	10.101	19.9
Total	50.760	100.0

Fonte: Tribunal Regional Eleitoral.

O parâmetro populacional a ser analisado é a proporção de intenção de votos. Portanto, o número de elementos a serem amostrados está num nível de 95% de confiabilidade e com a admissão de um erro de 5%, se a amostra fosse probabilística, seria de 384. Com a cpf, seria de 381 elementos (veja o tópico 3.4). Assim sendo, o erro em torno do valor central será menor que 5% se a amostra for de 500 elementos. O número de elementos a serem amostrados por região da cidade de Ibirité é dado pela tabela 02.

TABELA 02: Número de Elementos Para a Amostra por Região - 1998

Regiões	%	Nº de eleitores para a amostra
1	19.4	97
2 A	6.1	31
2 B	10.9	55
3	5.7	30
4	2.8	14
5	28.4	142
6	6.8	34
7	19.9	100
Total	100.0	503

Fonte: Tribunal Regional Eleitoral.

4.3 AS ESTRATIFICAÇÕES PARA A SELEÇÃO DOS ELEITORES

A idéia inicial era a de se amostrarem 500 elementos, mas devido aos erros de aproximação, foram amostrados 503 elementos (eleitores). O problema agora é verificar quais variáveis influenciariam a opinião do eleitor. Num primeiro instante, para a primeira amostra, as variáveis sexo, idade e faixa etária foram as escolhidas. Sabe-se pelo Censo Demográfico (IBGE) que 47.6% da população de Ibirité são representados pelo sexo masculino e 52.4% pelo feminino. Assim, tornou-se necessário subdividir o número de elementos a serem retirados de cada região, proporcionalmente, pela característica sexo (veja tabela 03).

TABELA 03: Número de Elementos Para a Amostra Por Região e Sexo - 1998

Regiões	Nº de eleitores para a amostra	Sexo	
		Feminino	Masculino
1	97	51	46
2 A	31	16	15
2 B	55	29	26
3	30	15	15
4	14	7	7
5	142	74	68
6	34	18	16
7	100	52	48
Total	503	262	241

Fonte: TRE e IBGE.

A faixa etária também foi definida como fator capaz de afetar a intenção de votos. A estratificação por faixa etária serviu de base para a tabela 04.

Para as duas últimas amostras, a estratificação foi elaborada por nível de instrução. Desta forma, para a segunda e terceira amostras, os resultados da tabela 04 sofreram outra alteração. A variável grau de instrução foi incluída como fator capaz de influir nas intenções de votos. Utilizando-se a distribuição da população por nível de instrução (fornecida pelo IBGE), chegou-se à distribuição da segunda e terceira amostras por região, sexo, faixa etária e nível de instrução (tabela 05).

TABELA 04: Total de Elementos a Serem Amostrados por Região e Faixa Etária - 1998

Faixa Etária / Região	16 a 17	18 a 24	25 a 34	35 a 44	45 a 54	55 ou +	Total
1	Fem. 4 Masc. 4	Fem. 10 Masc. 9	Fem. 15 Masc. 14	Fem. 10 Masc. 10	Fem. 5 Masc. 5	Fem. 7 Masc. 4	Fem. 51 Masc. 46
2 A	Fem. 1 Masc. 1	Fem. 3 Masc. 3	Fem. 5 Masc. 5	Fem. 3 Masc. 3	Fem. 2 Masc. 2	Fem. 2 Masc. 1	Fem. 16 Masc. 15
2 B	Fem. 2 Masc. 2	Fem. 6 Masc. 5	Fem. 9 Masc. 8	Fem. 6 Masc. 6	Fem. 3 Masc. 3	Fem. 3 Masc. 2	Fem. 29 Masc. 26
3	Fem. 1 Masc. 1	Fem. 3 Masc. 3	Fem. 3 Masc. 5	Fem. 3 Masc. 3	Fem. 3 Masc. 2	Fem. 2 Masc. 1	Fem. 15 Masc. 15
4	Fem. 1 Masc. 1	Fem. 2 Masc. 1	Fem. 1 Masc. 1	Fem. 1 Masc. 2	Fem. 1 Masc. 1	Fem. 1 Masc. 1	Fem. 7 Masc. 7
5	Fem. 7 Masc. 6	Fem. 15 Masc. 14	Fem. 20 Masc. 19	Fem. 16 Masc. 15	Fem. 8 Masc. 8	Fem. 8 Masc. 6	Fem. 74 Masc. 68
6	Fem. 1 Masc. 1	Fem. 4 Masc. 3	Fem. 5 Masc. 4	Fem. 4 Masc. 4	Fem. 2 Masc. 2	Fem. 2 Masc. 2	Fem. 18 Masc. 16
7	Fem. 4 Masc. 3	Fem. 10 Masc. 9	Fem. 15 Masc. 14	Fem. 11 Masc. 11	Fem. 6 Masc. 5	Fem. 6 Masc. 6	Fem. 52 Masc. 48
Total	40	100	143	108	58	54	**Geral** 503

Fonte: EARLR/1998

TABELA 05: Total de Elementos a Serem Amostrados por Região, Faixa etária e Nível de Instrução[9] - 1998

Faixa Etária / Região	16 a 17	18 a 24	25 a 34	35 a 44	45 a 54	55 ou +	Total
1	Fem. A- D-1 B- E- C- 3 F- Masc. A- D-1 B- 1 E- C- 2 F-	Fem. A-1 D-2 B- 2 E-1 C- 4 F- Masc. A- 1 D-2 B- 1 E- C-5 F-	Fem. A-1 D-2 B-4 E-1 C-7 F- Masc. A- 1 D-2 B- 3 E-1 C- 7 F-	Fem. A-1 D-1 B-3 E-1 C-4 F- Masc. A-1 D-1 B-3 E-1 C-4 F-	Fem. A-1 D- B-2 E- C-2 F- Masc. A-1 D- B-2 E- C-2 F-	Fem. A-3 D- B-3 E- C-1 F- Masc. A-2 D- B-1 E- C-1 F-	Fem. 51 Masc. 46
2A	Fem. A- D- B- E- C-1 F- Masc. A- D- B- E- C-1 F-	Fem. A- D-1 B-1 E- C-1 F- Masc. A- D-1 B-1 E- C-1 F-	Fem. A-1 D-1 B-1 E-1 C-1 F- Masc. A- D-1 B-1 E-1 C-2 F-	Fem. A-1 D- B-1 E- C-1 F- Masc. A- D-1 B-1 E- C-1 F-	Fem. A- D- B-1 E- C-1 F- Masc. A- D- B-1 E- C-1 F-	Fem. A-1 D- B-1 E- C- F- Masc. A- D- B-1 E- C- F-	Fem. 16 Masc. 15
2B	Fem. A- D-1 B- E- C-1 F- Masc. A- D- B-1 E- C-1 F-	Fem. A- D-2 B-1 E-1 C-2 F- Masc. A- D-1 B-1 E- C-3 F-	Fem. A-1 D-1 B-2 E-1 C-4 F- Masc. A- D-2 B-2 E-1 C-3 F-	Fem. A-1 D-1 B-1 E-1 C-2 F- Masc. A-1 D-1 B-1 E-1 C-2 F-	Fem. A-1 D- B-1 E- C-1 F- Masc. A-1 D- B-1 E- C-1 F-	Fem. A-1 D- B-1 E- C-1 F- Masc. A-1 D- B-1 E- C- F-	Fem. 29 Masc. 26
3	Fem. A- D- B- E- C-1 F- Masc. A- D- B- E- C-1 F-	Fem. A- D-1 B-1 E- C-1 F- Masc. A- D-1 B-1 E- C-1 F-	Fem. A- D-1 B-1 E- C-1 F- Masc. A-1 D-1 B-1 E- C-2 F-	Fem. A-1 D- B-1 E- C-1 F- Masc. A-1 D- B-1 E- C-1 F-	Fem. A- D-1 B-1 E- C-1 F- Masc. A- D- B-1 E- C-1 F-	Fem. A-1 D- B-1 E- C- F- Masc. A- D- B-1 E- C- F-	Fem. 15 Masc. 15
4	Fem. A- D- B- E- C-1 F- Masc. A- D- B- E- C-1 F-	Fem. A- D-1 B- E- C-1 F- Masc. A- D- B- E- C-1 F-	Fem. A- D- B-1 E- C- F- Masc. A- D- B- E- C-1 F-	Fem. A- D- B-1 E- C- F- Masc. A- D- B-1 E- C-1 F-	Fem. A- D- B- E- C-1 F- Masc. A- D- B- E- C-1 F-	Fem. A-1 D- B- E- C- F- Masc. A-1 D- B- E- C- F-	Fem. 7 Masc. 7
5	Fem. A- D-2 B-2 E- C-3 F- Masc. A- D-2 B-1 E- C-3 F-	Fem. A-1 D-3 B-2 E-2 C-5 F-2 Masc. A- D-3 B-2 E-2 C-5 F-2	Fem. A-2 D-3 B-4 E-3 C-7 F-1 Masc. A-2 D-3 B-3 E-2 C-8 F-1	Fem. A-2 D-2 B-4 E-2 C-6 F- Masc. A-2 D-2 B-3 E-2 C-6 F-	Fem. A-1 D- B-4 E- C-3 F- Masc. A-2 D- B-3 E- C-3 F-	Fem. A-4 D-1 B-2 E- C-1 F- Masc. A-2 D-1 B-2 E- C-1 F-	Fem. 74 Masc. 68

[9] Simbologia: A: Sem Instrução; B: 1 a 3 anos de estudo; C: 4 a 7 anos de estudo; D: 8 a 10 anos de estudo; E: 11 a 14 anos de estudo; F: 15 anos de estudo ou mais.

Continuação da Tabela 05

Faixa Etária / Região	16 a 17	18 a 24	25 a 34	35 a 44	45 a 54	55 ou +	Total
6	Fem. A- D- B- E- C-1 F- Masc. A- D- B- E- C-1 F-	Fem. A- D-1 B-1 E- C-2 F- Masc. A- D-1 B-1 E- C-1 F-	Fem. A- D-1 B-1 E-1 C-2 F- Masc. A- D-1 B-1 E- C-2 F-	Fem. A-1 D-1 B-1 E- C-1 F- Masc. A-1 D-1 B-1 E- C-1 F-	Fem. A- D- B-1 E- C-1 F- Masc. A- D- B-1 E- C-1 F-	Fem. A-1 D- B-1 E- C- F- Masc. A-1 D- B-1 E- C- F-	Fem. 18 Masc. 16
7	Fem. A- D-1 B-1 E- C-2 F- Masc. A- D-1 B-1 E- C-1 F-	Fem. A-1 D-1 B-2 E-1 C-5 F- Masc. A-1 D-1 B-4 E-1 C-2 F-	Fem. A-1 D-3 B-3 E-2 C-6 F- Masc. A-1 D-3 B-3 E-2 C-5 F-	Fem. A-2 D-1 B-3 E- C-5 F- Masc. A-2 D-1 B-3 E-1 C-4 F-	Fem. A-1 D-1 B-2 E- C-2 F- Masc. A-1 D-1 B-1 E- C-2 F-	Fem. A-2 D-1 B-2 E- C-1 F- Masc. A-2 D- B-2 E- C-2 F-	Fem. 52 Masc. 48
Total	41	100	145	108	58	54	**Geral** 503

Fonte: EARLR/1998.

4.4 PROCEDIMENTOS PARA SE RETIRAR AS AMOSTRAS E SEUS RESPECTIVOS ELEMENTOS DO ESPAÇO GEOGRÁFICO

Já estão definidos quantos elementos, com características predeterminadas, serão retirados para a composição das amostras. O problema principal, agora, é como retirar esses elementos da população, quando se tem apenas o traçado urbano da cidade.

A Cidade de Ibirité pode ser definida com a divisão das oito (8) regiões, segundo o TRE (veja figura 19). Uma solução para a retirada dos elementos da população seria dividir as regiões em áreas menores (conglomerados) e, por meio de um sorteio aleatório, escolher os conglomerados a serem amostrados. Isso foi feito e pode ser verificado pela figura 20**,** que mostra a região 2A com suas respectivas subdivisões (conglomerados).

As subdivisões poderiam ter sido feitas, utilizando-se o sistema de grade. Isso não foi feito, porque não existe simetria perfeita entre os quarteirões e poderia ocorrer sobreposição de áreas ou conglomerados na aplicação dos questionários.

Outro fato a ser observado é que, geralmente, neste tipo de amostragem em campo, para que realmente se captem as tendências populacionais, os entrevistadores, sempre que possível, têm de aplicar questionários em mais de uma região, de tal forma que a amostra fique bem distribuída espacialmente. Desse modo, os conglomerados são escolhidos aleatoriamente por região e entrevistador. Isto é, quanto maior o número de questionários a serem aplicados pelo entrevistador numa determinada região, maior o número de conglomerados sorteados para o entrevistador.

Geralmente, quando se repete a seleção de amostragem em determinada cidade, faz-se a permuta dos entrevistadores. O entrevistador nunca vai às mesmas regiões ou áreas. O coordenador de campo é o responsável pelo cumprimento da seleção.

O entrevistador, ao chegar ao campo, visita apenas 20% das casas do quarteirão. Para cada cinco casas, apenas uma é visitada (visita-se uma casa e pulam-se quatro). Ele visita todos os quarteirões do primeiro conglomerado. Se o entrevistador não cumpriu a sua quota, ele passa para o segundo conglomerado, sorteado aleatoriamente e antecipadamente, e assim sucessivamente, até que toda a sua quota seja cumprida. Geralmente, o entrevistador nunca cumpre a quota com a visita aos primeiros conglomerados. Quanto maior o número de estratificações, mais conglomerados ele terá de visitar.

As três seleções de amostragem não foram feitas simultaneamente, ocorreram respectivamente na terceira semana de agosto, primeira e terceira semanas de setembro (1998). Esse é um fato que pode levar a uma mudança de tendências, dependendo das estratégias políticas dos candidatos.

4.5 ALGUNS PROBLEMAS OCORRIDOS EM CAMPO

Quando se trata de trabalho em campo, por mais treinados que estejam os entrevistadores, diversos fatos podem provocar erros, alterando os resultados, principalmente se não houver um bom gerenciamento no campo. Esse trabalho de gerenciamento, é feito pelo coordenador de campo, que deve ter a sensatez como característica principal.

Neste estudo, alguns problemas foram detectados no campo:

i) grande parte dos eleitores de 16 e 17 anos optou por não votar, não providenciando o título em tempo;

ii) na fronteira da cidade de Ibirité com Belo Horizonte, a maioria dos eleitores, apesar de morar em Ibirité, possui título de Belo Horizonte;

iii) na 2ª e 3ª amostras, a dificuldade foi encontrar eleitores com nível de instrução mais elevado, porque Ibirité é uma cidade cuja população possui baixa renda e nível de instrução elementar.

A grande maioria dos problemas foi resolvida por meio da troca de áreas (conglomerados). Como o entrevistador tinha uma lista com uma seqüência de conglomerados, ele se dirigia, por ordem, para o conglomerado seguinte.

4.6 OS RESULTADOS DAS AMOSTRAS E DO TRE

Nesse estudo, os cargos considerados foram para Deputado Estadual, Deputado Federal, Governador e Presidente. As respostas para intenção de votos foram colocadas sob duas formas: resposta espontânea (O) e estimulada (E) para as três amostras. Para o cargo de Presidente, nas duas últimas amostras, o resultado fornecido foi o estimulado (E). As tabelas 06 a 09 fornecem os resultados (tendências de votos dos eleitores) das três amostras simultaneamente, além do resultado final obtido pelo TRE. Observa-se que não foram levados em conta todos os candidatos oficiais do TRE, mas apenas os candidatos relacionados com as respostas dos eleitores nas amostras. O resultado oficial do TRE foi colocado de duas maneiras: (C), com relação ao comparecimento e (P), com relação à população de 50.760 eleitores da cidade de Ibirité.

TABELA 06: Tendências e Votos Para Deputado Estadual - Cidade de Ibirité - 1998

Deputado Estadual	1ª Amostra %		2ª Amostra %		3ª Amostra %		TRE %	
	O	E	O	E	O	E	C	P
Dinis Antônio Pinheiro	19.3	38.2	15.7	33.0	19.8	37.1	26.9	22.7
Paulo Telles da Silva	8.0	23.1	11.2	24.2	13.5	21.4	15.1	12.7
Álvaro A. Teixeira Dias	1.8	7.6	1.2	5.0	0.8	3.5	1.2	1.0
Antônio G. de Oliveira	3.2	0.0	0.0	0.0	0.0	0.0	2.5	2.1
Outros	6.4	7.0	5.9	4.3	5.1	4.1	40.9	34.8
Indeciso	4.2	17.1	50.4	21.6	31.2	18.4	....	......
Anular ou não vai votar	4.0	4.2	3.8	6.2	2.4	6.1	13.4	26.7
NS/NR	17.1	2.8	11.9	5.7	27.3	9.4	....	......

Fonte: TRE e Pólis Consultoria.

TABELA 07: Tendências e Votos Para Deputado Federal - Ibirité - 1998

Deputado Federal	1ª Amostra %		2ª Amostra %		3ª Amostra %		TRE %	
	O	E	O	E	O	E	C	P
Eliseu Resende	8.8	28.9	5.9	23.3	5.5	16.9	11.6	9.9
Vittório Medioli	8.6	18.3	9.0	16.6	4.7	13.5	6.6	5.6
Maria do C. L. Perpétuo	0.0	0.0	0.0	0.0	0.0	0.0	7.4	6.3
Outros	4.6	3.4	9.7	9.7	10.0	9.0	60.2	50.8
Indeciso	46.8	24.1	$\}$ 53.7	$\}$ 38.8	45.7	34.1	......	......
Anular ou não vai votar	5.4	18.5			0.2	9.2	14.2	27.4
NS/NR	25.9	6.8	21.6	11.6	33.9	17.3	......	......

Fonte: TRE e Pólis Consultoria.

TABELA 08: Tendências e Votos Para Governador - Ibirité -1998

Governador	1ª Amostra %		2ª Amostra %		3ª Amostra %		TRE %	
	O	E	O	E	O	E	C	P
Itamar A. C. Franco	21.1	36.9	26.6	35.9	26.5	31.6	34.3	29.9
Eduardo B. de Azeredo	16.1	31.1	21.4	31.8	22.7	31.2	23.5	19.9
Patrus A. de Sousa	6.6	11.2	10.7	13.8	9.4	12.4	17.8	15.1
Outros	2.4	1.0	1.0	0.5	0.6	1.6	0.5	1.3
Indeciso	31.7	13.7	26.8	10.9	20.0	14.5	......	......
Anular ou não vai votar	3.4	4.0	2.1	3.6	4.1	3.3	23.9	33.8
NS/NR	18.7	2.2	11.4	3.3	16.7	5.4	......	......

Fonte: TRE e Pólis Consultoria.

TABELA 09: Tendências e Votos Para Presidente - Ibirité - 1998

Presidente	2ª Amostra %	3ª Amostra %	TRE %	
	E	E	C	P
Fernando H. Cardoso	34.0	35.1	34.8	29.4
Luiz I. Lula da Silva	32.8	31.4	33.4	28.3
Ciro Ferreira Gomes	4.5	4.7	7.1	6.0
Enéas Ferreira Carneiro	7.1	4.5	1.9	1.6
Outros	0.5	0.4	1.1	0.9
Indeciso	13.5	15.7	......	......
Anular ou não vai votar	4.8	3.9	21.7	33.8
NS/NR	2.9	4.3	......	......

Fonte: TRE e Pólis Consultoria.

De um total de 50.760 eleitores, 42.961 compareceram às urnas, perfazendo um total de 15.36% de abstenções. Na coluna dos resultados oficiais do TRE, a opção C representa os votos brancos ou nulos e a opção P, os votos brancos, nulos e abstenção.

4.7 COMPARAÇÃO DOS RESULTADOS UTILIZANDO A INFERÊNCIA ESTATÍSTICA

A proporção populacional (teoricamente) segue a distribuição hipergeométrica (amostra sem reposição), mas, para pequenas amostras, podem-se aproximar os resultados pela distribuição Binomial (veja anexo 02). Assim, a estatística para os testes será fornecida pela quantidade $Z = (\hat{p} - p)/\sqrt{pq/n}$, e o módulo deste valor será comparado com o valor 1.96 da distribuição Normal, o que implica um nível de 5% de significância ou 95% de confiabilidade.

4.7.1 Nomenclatura

p proporção populacional (resultado oficial com relação à população).

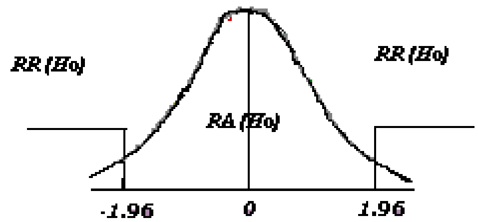
$\hat{p}$ proporção da amostra relacionada com o candidato.

$q = (1 - p)$ proporção populacional não relacionada com o candidato.

n número de elementos da amostra.

Z valor da distribuição Normal (95% ou 100(1-α)% de confiabilidade ou 5% de significância).

O teste é composto de duas hipóteses: a primeira é chamada de hipótese nula



($H_0 : p =$Resultado do TRE) e a segunda, denominada por hipótese alternativa[10] ($H_1 : p \neq$ Resultado do TRE), assim, se o valor calculado do módulo de Z for maior que 1.96, a amostra não pertencerá à população (diferença significante). Caso contrário, o resultado amostral será dito igual ao resultado populacional (diferença insignificante).

Se a diferença entre o valor da amostra e o valor da população for suficientemente grande, o valor de Z (calculado) estará situado na região de rejeição, e isto representa a não conformidade da amostra com a população ou que a amostra não faz parte da população.

Lembrando o fato de que as seleções de amostragem foram feitas bem antes das eleições, e que a última seleção de amostragem (terceira) foi feita há quinze dias do pleito, pode ter ocorrido uma migração de votos, que dependeria das estratégias políticas dos candidatos. A tabela 10 fornece os valores calculados da estatística z por candidato.

Ao se comparar a intenção de votos da terceira amostra (espontânea) com os resultados oficiais (relacionados com a população total), apenas dois resultados são significantes (Eliseu Resende e Fernando H. Cardoso), isto é, a intenção de votos obtida por meio da terceira amostra não é compatível com os resultados populacionais. Mas, se a comparação for feita com o resultado oficial dos eleitores que compareceram às urnas, para presidente, o resultado se torna insignificante para Fernando H. Cardoso (Z = 0.768), e significante para Ciro Gomes (Z = 2.096) e, para deputado federal, o resultado se torna insignificante ao ser comparado com o resultado da primeira amostra ($\hat{p} = 0.088$ e Z = -0.826). Portanto, pode-se dizer, para uma análise geral, que a terceira amostra, apesar de ter sido obtida há quinze (15) dias do pleito, representou muito bem a população. Algumas tendências de aumento ou decrescimento de votos, da primeira até a terceira amostra, também podem ser confirmadas com o resultado final do pleito.

Outro fato a ser observado é que, da primeira para a segunda seleção de amostragem, houve modificação da seleção de amostragem. Isso também influenciou na coleta dos dados, devido ao tempo muito maior que os entrevistadores levaram para que suas respectivas quotas fossem cumpridas.

TABELA 10: Valores da Estatística Z para os Primeiros Candidatos, por Cargo

Deputado Estadual	**3ª Amostra (O)**	**TRE (P)**	**Z**
Dinis A. Pinheiro	19.8	22.7	-1.550
Paulo Telles da Silva	13.5	12.7	0.537
Álvaro A. Teixeira Dias	0.8	1.0	-0.450
Deputado Federal	**3ª Amostra (O)**	**TRE (P)**	**Z**
Eliseu Rezende	5.5	9.9	-3.300
Vittório Medioli	4.7	5.6	-0.876
Governador	**3ª Amostra (O)**	**TRE (P)**	**Z**
Itamar A. C. Franco	26.5	29.9	-1.665
Eduardo B. de Azeredo	22.7	19.9	1.689
Patrus A. de Sousa	9.4	15.1	-0.970
Presidente	**3ª Amostra (E)**	**TRE (P)**	**Z**
Fernando H. Cardoso	35.1	29.4	2.806
Luiz I. Lula da Silva	31.4	28.3	1.543
Ciro Ferreira Gomes	4.7	6.0	-1.288

Fonte: EARLR / 1999.

[10] Triola, 1999.

Outra consideração relevante está no fato de que apesar de a estratificação na seleção das amostras não ter levado em conta as variáveis renda e religião, os resultados foram compatíveis com os resultados populacionais.

4.8 - SUGESTÕES DE COMO MELHORAR A SELEÇÃO DE AMOSTRAGEM

Nesse estudo de caso, nota-se claramente que a seleção de amostragem foi satisfatória, pois, na amostra, os candidatos mais votados mantiveram a mesma posição dos resultados populacionais. Entretanto, se for levada em conta a posição do candidato por região eleitoral, de tal forma que o candidato tenha condições de atuar e melhorar seu posicionamento dentro da região eleitoral, cada região eleitoral pode ser vista como uma subpopulação, onde o tratamento será individualizado. Ibirité teria, então, (8) oito subpopulações e o tamanho da amostra seria calculado individualmente para cada subpopulação. A tabela 11 mostra os valores da amostra por subpopulação (veja anexo 04). Observa-se que se existissem valores de p ou uma estimativa $\hat{p}$ por subpopulação, o tamanho da amostra, na maioria dos casos, seria menor.

Neste caso, as subpopulações podem ser analisadas independentemente uma da outra, e as atuações podem ser verificadas espacialmente. Sugere-se inicialmente a retirada de uma amostra piloto para verificação das tendências, de tal modo que a amostra por subpopulação não tenha um tamanho superestimado (p=0.50).

TABELA 11: Distribuição da Amostra Por Região Eleitoral

Região	Número de elementos por região eleitoral, nível de 95% de confiabilidade e erro máximo de 5%					
	Sem CPF P =0.50	Com CPF P=0.50	TRE:P máximo = 0.30		Amostra: P máximo = 0.35	
			Sem CPF	**Com CPF**	**Sem CPF**	**Com CPF**
1	384	370	323	313	350	339
2A	384	342	323	293	350	315
2B	384	359	323	306	350	330
3	384	339	323	291	350	331
4	384	302	323	263	350	281
5	384	374	323	316	350	342
6	384	347	323	295	350	318
7	384	370	323	313	350	339
Total	**3072**	**2802**	**2584**	**2390**	**2800**	**2595**

Fonte: EARLR / 1999.

Para exemplificar, supõe-se que para uma amostra piloto, obteve-se os seguintes resultados para as subpopulações (regiões eleitorais) da cidade de Ibirité, para o candidato X e que com estes resultados outra amostra foi selecionada (veja tabela 12):

TABELA 12: Resultados Hipotéticos a Partir de uma Amostra Piloto com Cálculo de Tamanho de Amostra e Intervalos de Confiança.

Regiões	$\hat{p}_i$ **(proporção da amostra piloto por região)**	N_i **(total de eleitores por região)**	n_i **(tamanho estimado da amostra por região), sem CPF**	$\hat{p}_i$ **(nova proporção utilizando-se o tamanho estimado para a amostra)**	**IC de 95% Para a proporção populacional de cada subpopulação**
1	0.10	9848	138	0.11	[0.06 ; 0.16]
2A	0.05	3096	73	0.04	[0.00; 0.09]
2B	0.15	5533	196	0.17	[0.12 ; 0.22]
3	0.20	2893	246	0.22	[0.17; 0.27]
4	0.25	1421	288	0.24	[0.19 ; 0.29]
5	0.05	14416	73	0.07	[0.02 ; 0.12]
6	0.10	3452	138	0.12	[0.07 ; 0.17]
7	0.30	10101	323	0.33	[0.28 ; 0.38]
Total	....	50760	1475	....	....

Fonte: EARLR / 1999.

Em média, o candidato terá 8183 votos, que correspondem a $\sum N_i \hat{p}_i$. Se o candidato necessitar de 10000 votos para ganhar as eleições, na pior das hipóteses ele terá 5255 votos (utilizando-se o limite inferior dos intervalos). Com relação à população total (50760 eleitores), o candidato necessitaria em torno de 20% dos votos. Assim, poderemos espacializar as regiões críticas para o candidato em questão (veja figura 21). Observa-se que a estimativa de n_i foi baseada em um erro de 0.05 em torno do valor central.

As seleções de amostragem feitas no estudo de caso, apesar de satisfatórias, não permitem planejamento e atuações individuais por regiões eleitorais, mas o interesse estava apenas na intenção de votos de todo eleitorado da cidade (estimativa por ponto no espaço). Mantendo-se o mesmo nível de confiabilidade e erro, pode-se observar que o tamanho da amostra aumenta substancialmente, mas os benefícios são inegáveis quando se trabalha com as subpopulações. A seleção da amostra por região eleitoral permite atuações concretas e diretas, minimizando-se os custos, além de fornecer melhores opções nas aplicações.

A seleção da amostra por região eleitoral deve seguir os mesmos procedimentos do estudo de caso, isto é, a seleção de amostragem será por quotas, e as estratificações começam pela subpopulação. Neste caso, as amostras também não serão aleatórias, pois não se conhece a distribuição real dos eleitores pelas variáveis de controle. Pela tabela 12, o número de elementos para a amostra seria de 246, mas devido às aproximações, o número de elementos para a amostra, depois das estratificações, passou para 247 elementos (veja tabela 13).

Para todo o desenvolvimento deste trabalho, não se utilizou o sistema de coordenadas para a seleção direta das amostras (veja tópico 3.2), devido aos seguintes motivos:

a) suposição de que diversas variáveis (sexo, faixa-etária e nível de instrução) influenciavam a opinião do eleitor, além das outras variáveis (religião e renda) com distribuições populacionais desconhecidas;
b) a amostragem de mapas supõe algumas características conhecidas (mapeadas);
c) a amostragem por quotas seria inviável para a escolha de pontos, utilizando-se o traçado urbano da cidade e o sistema de coordenadas, pois, com a utilização dos

projetos clássicos de amostragem, todos os supostos conglomerados poderiam ser visitados pelo entrevistador, e pela escolha de pontos coordenados não se saberia com antecedência quantos pontos deveriam ser escolhidos de modo que a seleção fosse satisfeita;

d) os problemas que ocorrem em campo têm melhor solução com o sistema de seleção utilizado;

e) a seleção pelo sistema de coordenadas supõe aleatoriedade em todo o procedimento, e apenas no caso em que o eleitor estiver georeferenciado é que seria viável a seleção pelo sistema de coordenadas, como também a utilização ampla dos Sistemas de Informações Geográficas - um mapa de pontos para cada quadrícula da distribuição da amostra, representando os eleitores da população com as características comuns e;

f) a população é discretamente distribuída, portanto, podem ser usados os projetos clássicos de amostragem.

TABELA 13: Distribuição da Amostra por Sexo, Faixa-etária e Nível de Instrução[11]

Para a Região Eleitoral 3 com n = 247

Faixa Etária / Região 3	16 a 17	18 a 24	25 a 34	35 a 44	45 a 54	55 ou +	Total
Sexo Feminino	A = 0 B = 1 C = 6 D = 2 E = 0 F = 0	A = 1 B = 3 C =14 D = 6 E = 3 F = 0	A = 2 B = 6 C =21 D = 6 E = 4 F = 0	A = 3 B = 7 C = 13 D = 2 E = 1 F = 0	A = 4 B = 5 C = 5 D = 1 E = 0 F = 0	A = 7 B = 4 C = 2 D = 0 E = 0 F = 0	**17** **26** **61** **17** **8** **0**
Sub - total	**9**	**27**	**39**	**26**	**15**	**13**	**129**
Sexo Masculino	A = 0 B = 2 C = 5 D = 2 E = 0 F = 0	A = 1 B = 3 C =14 D = 5 E = 1 F = 0	A = 2 B = 6 C =18 D = 6 E = 3 F = 0	A = 2 B = 7 C =13 D = 3 E = 1 F = 0	A = 2 B = 5 C = 5 D = 1 E = 0 F = 0	A = 4 B = 4 C = 3 D = 0 E = 0 F = 0	**11** **27** **58** **17** **5** **0**
Sub - total	**9**	**24**	**35**	**26**	**13**	**11**	**118**
Total Geral	**18**	**51**	**74**	**52**	**28**	**24**	**247**

Fonte: EARLR / 1999.

11 **Simbologia**: A: Sem Instrução; B: 1 a 3 anos de estudo; C: 4 a 7 anos de estudo; D: 8 a 10 anos de estudo; E: 11 a 14 anos de estudo e; F: 15 anos de estudo ou mais.

5. CONCLUSÕES E AVALIAÇÕES

No estudo de caso na cidade de Ibirité, apesar de uma localização espacial imperfeita (do eleitor), conseguem-se melhores resultados do que em pesquisas "ditas aleatórias". Por outro lado, muitas das pesquisas de opinião "aleatórias" não levam em conta as características do eleitor (sexo, faixa-etária, nível de instrução e outras). O procedimento, apesar de totalmente aleatorizado, fornecerá estimativas tendenciosas e poderá levar ao descrédito das aplicações.

Na seleção por quotas, a retirada do elemento "eleitor" não é aleatória devido ao não georeferenciamento, entretanto, as regiões eleitorais foram espacializadas a partir do traçado urbano do município, o que favoreceu amplamente no planejamento e execução do experimento.

No planejamento do experimento, a suposição de que as regiões eleitorais tivessem equivalentemente eleitores de todas as faixas etárias, sexo e nível de instrução foi utilizada. Tal suposição, em muitos casos, não condiz com a realidade, dificultando a execução do experimento. Para o estudo de caso de Ibirité, os resultados foram satisfatórios, tanto sociologicamente (comportamento político dos eleitores), quanto estatisticamente (estudo utilizando a inferência estatística), estando bem próximos da realidade. O ideal seria trabalhar individualmente com cada região eleitoral, o que permitiria, além de um enfoque mais amplo, aplicações orientadas.

A seleção da amostra por região eleitoral fornece aplicações e comparações amplas, tratando-se, portanto, da melhor maneira de análise no espaço.

A partir do momento em que os eleitores estiverem na forma georeferenciada, o estudo será mais amplo, pois permitirá a utilização dos Sistemas de Informações Geográficas, além de uma seleção aleatória dos eleitores.

Se os eleitores estivessem representados na estrutura vetorial (Abreu, 1995), poderiam ser distribuídos em mapas que os representariam por características similares, e desta forma, a amostragem poderia ser totalmente aleatorizada e espacializada.

A estrutura vetorial é representada por um conjunto de coordenadas, no qual um conjunto de pontos (eleitores) representaria os espaços reais. Assim, seria possível distribuir os eleitores por características similares, em que cada eleitor assumiria a forma vetorial (veja tabela 14). Todos os eleitores seriam representados por uma matriz, conforme tabela 15. Assim, grupos similares de eleitores poderiam ser agrupados em matrizes individuais, que poderiam ser transformadas em mapas (pontuais) e a seleção amostral seria aleatória por pontos. Observa-se que para cada variável inserida no estudo, a matriz aumenta em uma coluna, e com as informações utilizadas no estudo de caso, (576) quinhentos e setenta e seis mapas, representando os eleitores com características similares, poderiam ser confeccionados, isto sem se levar em conta as variáveis renda e a religião.

Os dados, quando espacializados, fornecem uma quantidade muito maior de informações, possibilitando o uso de técnicas mais refinadas para análises e comparações dos mesmos. Entretanto, observa-se que a matriz com os dados dos eleitores conteria (355.320) trezentos e cinqüenta e cinco mil e trezentos e vinte informações, e que para cada variável inserida no estudo mais (50.760) cinqüenta mil setecentos e sessenta informações seriam adicionadas à matriz, com o aumento de uma coluna. Esta base de dados ficaria muito cara e deveria ser sempre atualizada.

TABELA 14: Estrutura Vetorial Para Um Eleitor[12].

Características	Variáveis
Localização (coordenadas)	$X_{i,1}$ e $X_{i,2}$
Região Eleitoral	$X_{i,3}$
Sexo	$X_{1,4}$
Faixa-etária	$X_{i,5}$
Nível de Instrução	$X_{i,6}$

Vetor eleitor $_i$ (X_{i1}, X_{i2}, X_{i3}, ...,X_{i6})

TABELA 15: Estrutura Matricial Para Todos os Eleitores do Município [13].

$$
\begin{bmatrix}
X_{(1,1)} & X_{(1,2)} & X_{(1,3)} & X_{(1,4)} & X_{(1,5)} & X_{(1,6)} \\
X_{(2,1)} & X_{(2,2)} & X_{(2,3)} & X_{(2,4)} & X_{(2,5)} & X_{(2,6)} \\
X_{(3,1)} & X_{(3,2)} & X_{(3,3)} & X_{(3,4)} & X_{(3,5)} & X_{(3,6)} \\
\ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\
\ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\
\ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\
X_{(50760,\ 1)} & X_{(50760,\ 2)} & X_{(50760,\ 3)} & X_{(50760,\ 4)} & X_{(50760,\ 5)} & X_{(50760,\ 6)}
\end{bmatrix}
$$

[12] $X_{i,1}$ = abscissa (localização) , $X_{i,2}$ = ordenada (localização), $X_{i,3}$ = Região Eleitoral, $X_{1,4}$ = Sexo, $X_{i,5}$ = Faixa-etária e $X_{i,6}$ = Nível de Instrução.

[13] Cada linha representa um eleitor e cada coluna as variáveis que influem na opinião do eleitor. N =50760 = número de eleitores do município de Ibirité.

Uma solução, para se construir a base de dados, estaria no acúmulo de informações, isto é, para cada pesquisa agrupariam-se os dados não cadastrados aos cadastrados, e atualizariam-se os dados já cadastrados, desta forma, em alguns anos de estudo já se teria uma base respeitável de informações. Assim, as estratificações seriam feitas de acordo com a distribuição das variáveis por região eleitoral, e os eleitores poderiam ser escolhidos aleatoriamente.

6. ABSTRACT

The purpose of this work is making a search on spatial sampling, starting from a classical pattern of probabilities, showing as well the importance, the primacy of non probabilistic selections, as one deals with voter's opinion. It also aims at illustrating the theoretical content by means of the study case for the Ibirité town - Minas Gerais - a combined sampling selection in which one discusses not only the validity of the sampling selection, but also error and dimensioning, besides the spatial aspect of the variables.

These analyses will be important starting from the point one can state that either a selection through sampling has a smaller variability or that the selection leads to a sample that actually belongs to the population being studied, namely, that fulfills not only one speciality.

Understanding the distribuition of the population is of utmost importance for one to carry out the proposed analyses. In case of lack of such a knowledge, assumptions will be presented in the course of the work.

The study mentioned herein does not intend, at any moment, to comprehend the entirety. Ideas are searched for, which may bring about a better understanding of some sampling projects and as for the study of case, the reason of the higher quality among some projects.

Nevertheless, the important is the sampling project itself, that is, the several reasons capable of leading to an ideal sampling project which will bring satisfactory results and will the enhance the knowledge for better sampling selections in the future.

8. REFERÊNCIAS BIBLIOGRÁFICAS

[1] ABREU, João Francisco. **Manufatura Integrada Por Computador; contexto , tendências e técnicas**. Belo Horizonte: Editora Terra, L.D. CNPQ / SEBRAE/ CEFET/ Minas Gerais, 1995.

[2] ADLER, Max K. **A Moderna Pesquisa de Mercado**. 2.ed. São Paulo: Pioneira, 1971. 138p.

[3] AGUIAR, Marco Antônio de Souza. **Manual Básico de Pesquisa de Mercado**. Brasília: Sebrae, 1998. 111p.

[4] AMORIM FILHO, Oswaldo Bueno. **Reflexões Sobre as Tendências Teórico Metodológicas da Geografia**. Belo Horizonte: Instituto de Geo-Ciências UFMG. 1995. 56p.

[5] ANDRADE, Manoel Correa. **Geografia Ciência da Sociedade**: uma introdução à análise do pensamento geográfico. São Paulo: Atlas, 1987. 143p.

[6] BONINI, Edmundo Eboli; BONINI, Sérgio Eboli. **Estatística**: teoria e exercícios. São Paulo: L.P.M., 1972. 443p.

[7] BOTELHO, Caio Lóssio. **A Filosofia e o Processo Evolutivo da Geografia**. Fortaleza. Universidade Federal do Ceará, 1987.

[8] BOWKER, Albert H.; LIEBERMAN, Gerald J.. **Engineering Statistics.** 2 ed. New Jersey: Prentice-Hall,Inc., 1972. 641p.

[9] BURT, James E. ; BARBER, Gerald M.. **Elementaty Statistics for Geographers**. 2 ed. New York: The Guilford Press, 1996. p. 242-251

[10] COCHRAN, William Gemmmell. **Sampling Techniques**. New York: John Wiley & Sons, 1977. 428p.

[11] CONVERSE, E. Philip. The Nature of Belief Systems in Mass Publics. In: APTER, D.(Ed.). **Ideology na Discontent.** New York: The Free Press of Glencoe, 1964. p.206-261.

[12] CHRISTOFOLETTI, Antônio. **Perspectivas da Geografia.** 2.ed. São Paulo: Difel, 1985. 318p.

[13] DOLLFUS, Olivier. **O Espaço Geográfico**. 4.ed. São Paulo: Difel, 1982. 121p.

[14] ELEIÇÕES 1998. [online] Disponível na Internet www.tre.gov.br. Arquivo capturado em 07 de outubro 1998.

[15] CREIGHTON, J. H .C.. **A First Course in Probability Models And Statistical Inference**. 1.ed. New York: Springer-Verlag, 1994. 717p.

[16] KOHLER, Heinz Charles. **Geomorfologia Cárstica na Região de Lagoa Santa, Minas Gerais**. São Paulo: Departamento de Geografia da Universidade de São Paulo, 1989. 113p. (Tese, Doutorado em Geologia.)

[17] GERARDI, Lúcia Helena de Oliveira, SILVA, Bárbara Christine Maria Nentwig. **Quantificação em Geografia**. São Paulo: Difel, 1981. 161p.

[18] GOODEY, Brian; GOLD, Brian. **Geografia do Comportamento e da Percepção**. Belo Horizonte: Instituto de Geo-Ciências da UFMG, 1986. 49p.

[19] HAMMOND; CULLAGH MC P.S. **Quantitative Techniques in Geography An Introduction**. London: Oxford University Press, Ely House, 1974. p.108-133.

[20] IBIRITÈ. [online] Disponível na Internet via www.almg.gov.br/munmg/M298.htm. Arquivo capturado em 14 de Junho 1999.

[21] KUHN, S. Thomas. **As Estruturas das Revoluções Científicas**. São Paulo: Perspectiva , 1975. 262p.

[22] LOURENÇO FILHO, Rui de C. B. **Controle Estatístico de Qualidade**. Rio de Janeiro: Livros Técnicos e Científicos, 1980. 223p.

[23] LEVIN, Jack. **Elementary Statistics in Social Research.** 2ed. New York: Harper & Row Publishers, 1977. 310p.

[24] MAAKAROUN, Bertha – **Mídia e Decisão de Voto.** Belo Horizonte: Departamento de Ciência Política da UFMG, 1994. (Tese, Mestrado Em Ciências Políticas.)

[25] MEYER , Paul L. **Probabilidade**: aplicações à estatística. Rio de Janeiro: Livros Técnicos e Científicos, 1984. 426p.

[26] MILBRATH, Lester. **Political Participation**. Chicago: Rand Mcnally, 1965.

[27] NEUMAN, W. Russel. **The Paradox of Mass Politics, Knowledge and Opinion in the American Electorate**. Cambridge: Harvard University Press, 1986.

[28] OLSON, M. **The logic of collective Action: Publics Goods and the Theory of Groups**. Cambridge: Harvard University Press, 1965.

[29] PAIVA, Antônio Fabiano. **Estatística**. Belo Horizonte: Universidade Federal de Minas Gerais, 1981,2v..

[30] PIÃO, Antônio Carlos S., HENRY, Raoul. **Estudo de Caso**: Transporte de nitrogênio, fósforo e sedimentos pelo Ribeirão dos Carrapatos (Município de Itaí, SP); Análise Ambiental "estratégias e ações". São Paulo: Fundação Salim Farah Maluf, 1995. 381p.

[31] SANTOS, Milton. **Técnica Espaço Tempo**. Globalização e meio técnico científico informacional. 3.ed. São Paulo: Editora Hucitec, 1997. 190p.

[32] SANTOS, Milton. **A Natureza do Espaço**. São Paulo: Editora Hucitec, 1996. 308p.

[33] SILVA, Jorge Xavier; SOUZA, Marcelo José Lopes. **Análise Ambiental**. Rio de Janeiro: Editora UFRJ, 1987. 196p.

[34] SOARES, José Francisco; FARIAS, Alfredo Alves; CESAR, Cibele Comini. **Introdução À Estatística**. Rio de Janeiro: Guanabara Koogan S.A., 1991. 378p.

[35] TAGLIACARNE, G.. **Pesquisa de Mercado**: Técnica e prática. São Paulo: Atlas, 1986.

[36] TEIXEIRA, Amândio Luís de Almeida; CHRISTOFOLETTI, Antônio. **Sistemas de Informações Geográficas:** Dicionário Ilustrado. São Paulo: Hucitec, 1997. 224p.

[37] THOMPSON, K. Steven. **Sampling.** New York: John Wiley & Sons, 1992. 343p.

[38] TRIOLA, Mário F.. **Introdução À Estatística**. Rio de Janeiro: LTC, 1999. 410p.

[39] WALLIS, W. Allen , HARRY, Roberts V. **Curso de Estatística.** Rio de Janeiro: USAID, 1964, 2v..

[40] WERKEMA, Maria Cristina Catarino. **Ferramentas Estatísticas Básicas Para o Gerenciamento de Processos**. Belo Horizonte: Fundação Christiano Ottoni, 1995, 404p.

[41] WERKEMA, Maria Cristina Catarino. **Como Estabelecer Conclusões Com Confiança: Entendendo Inferência Estatística**. Belo Horizonte: Fundação Christiano Ottoni, 1996, 279p.

[42] YULE, George Udny.; KENDALL, G. Maurice. **Introduccion a la Estadística matemática**. 3.ed. Madrid: Aguilar, 1959. 740p.

ANEXO 01

TABELA DE NÚMEROS ALEATÓRIOS GERADA POR COMPUTADOR

1. A Distribuição Uniforme como geradora da tábua de números aleatórios (TNA), com a utilização do "software" Minitab

TABELA 16: Tábua de Números Aleatórios -TNA

Linha							
1	879181	275333	432174	685182	199873	795551	839551
2	453026	315157	005157	207439	385871	407274	370066
3	607256	571309	496023	512034	329431	240379	661850
4	863827	826283	188408	477339	752853	147212	364217
5	576976	216528	642930	440333	318188	205120	100769
6	620995	587744	234772	634124	259302	095339	708048
7	523758	568035	829292	237659	675675	647492	664292
8	763379	604066	881625	210386	280218	191288	642319
9	204813	844022	202342	606926	213188	634730	467904
10	595765	601481	128535	674930	375941	054557	800008
11	716417	892152	288333	544325	669583	538301	244392
12	662904	653347	633012	340227	398587	884038	198329
13	200875	302914	348114	558653	656369	249354	486560
14	123443	792963	770763	312079	467434	088203	881500
15	064741	130010	244675	108956	408481	035469	869797
16	265741	250067	589534	839301	873196	018950	248793
17	731206	160812	515530	720317	772256	434101	893444
18	544052	469794	031524	628037	376409	423368	170670
19	668157	765676	183212	220247	574438	107595	214073
20	223535	806256	793664	778093	693827	229776	880062
21	056595	732819	457259	787937	283026	356321	207668
22	643427	664729	475365	814929	756418	385024	475765
23	379597	353100	668405	013128	279955	487865	378761
24	877093	834239	167017	368033	510596	196532	395592
25	407072	243831	617663	408030	204487	258753	220285
26	722876	343520	597188	322037	458394	425390	102585
27	313882	180257	107158	392089	382208	628423	405873
28	538453	225472	619562	478439	729685	734884	307580
29	185906	725002	733468	262954	615308	180607	653942
30	570995	225850	519650	150755	588017	188039	220153
31	527785	797163	355248	865250	762726	062379	727674
32	373538	878412	293807	611835	528422	203320	082313
33	374958	591277	682677	159173	871361	153452	514538
34	872007	128734	636460	456386	468346	055975	345071
35	718657	129634	664194	061052	319371	395292	850662
36	894767	159791	103301	372063	799066	124560	398608
37	363284	196843	853602	165334	887054	268779	706724
38	034998	491746	359540	166148	892098	743274	053289
39	738475	450657	415422	349428	767497	798040	820590
40	723012	233648	474870	095320	080681	708820	828704
41	270855	315395	655885	497020	755538	364124	235176
42	478423	293465	189313	272763	797963	239458	878869
43	670195	331262	038609	339360	497162	229110	365428
44	628254	425230	753917	577987	888467	451891	800992
45	289266	042075	545389	795056	029648	424617	363456
46	154267	218502	378005	313011	550301	626088	680532
47	663289	342750	761088	874659	283277	414827	863117

Continuação da Tabela 16							
48	108333	175168	682202	187192	199934	815491	553965
49	611311	163967	784832	794250	235467	546838	052389
50	446226	584489	563431	490618	791621	754934	360336
51	431015	657071	247864	267616	744339	448410	467562
52	140738	049922	534726	572437	170020	591612	159934
53	649789	497088	475990	088476	133298	573131	177308
54	068901	821732	257123	312841	722226	622314	845607
55	783892	360805	678626	810452	731866	337663	081901
56	393543	862475	607803	097671	636810	456995	857790
57	453444	876811	463746	214139	793828	416427	269425
58	281507	411080	716773	409692	357678	283909	359439
59	857356	506543	524498	404625	840477	084844	439522
60	307557	280506	399087	238034	102755	590501	505439
61	719198	339175	192745	622226	595585	281346	785692
62	179492	871195	637465	284859	578208	245994	704817
63	343578	082244	110899	148506	851871	388471	415302
64	153431	131895	623209	768356	162097	319125	853233
65	136047	174281	576210	031919	693100	474146	870083
66	621271	620759	740142	737121	302969	151938	307245
67	736208	334540	580238	793072	786306	566841	272184
68	277919	893660	336271	894286	370864	453251	550491
69	240084	571007	539492	397594	028804	138125	237853
70	537404	494091	198787	361696	243443	041680	141624
71	031336	878586	278622	692185	071927	120619	659983
72	401960	377453	767324	432113	336766	295197	871898
73	589198	062312	672695	474233	235446	886023	349645
74	334043	309353	625704	898712	804944	389855	134339
75	181386	503031	267369	894869	860765	434286	332764
76	175596	070307	141065	853236	245910	671758	353875
77	770385	832835	862788	167030	214905	488492	408850
78	239161	715455	808966	844772	499644	841376	886246
79	048816	391972	256193	251445	838300	738341	881637
80	399103	315051	051113	427662	547599	475172	700276
81	661554	285687	877096	112032	615213	215484	754064
82	875931	277061	736373	136970	355493	203854	450936
83	156496	735788	124960	755416	023446	714106	229869
84	269708	164742	293250	650890	383475	746840	866799
85	719692	115285	720019	284730	177722	377456	568557
86	297628	707038	590314	468749	455419	706226	652633
87	081642	152709	900850	592966	605841	693030	287160
88	863266	195490	018000	188586	170362	188208	783488
89	017288	683124	734879	083440	576247	463245	678596
90	168418	305759	647898	646971	391172	182977	078134
91	390175	592998	887850	521952	017325	832370	609987
92	310272	712064	111906	723388	096209	633162	624685
93	098851	243349	072756	184370	820082	321500	162518
94	524441	216603	370742	349571	566190	393309	827131
95	758392	147638	855404	152263	384463	039093	764876
96	775586	830642	366915	538251	538476	066331	885832
97	679444	115009	700187	715241	542756	482549	242413
98	399522	857049	818773	893168	831871	734662	755324
99	778815	733370	183133	749420	359001	563911	341469
00	101246	773088	747201	320676	364377	721654	876470

Para gerar a TNA, utilizando a distribuição Uniforme, foram observados os seguintes passos:

1.1) Clicar com o mouse no comando Calc do menu principal.

1.2) Clicar com o mouse nos seguintes comandos: Random Data - Uniform - Generate rows of Data (digitar no quadrado o número de valores a serem gerados) - Store in column(s) (digitar no quadrado quantas colunas deseja) - Lower endpoint (digitar o valor do limite inferior do domínio da função) - Upper endpoint (digitar o valor do limite superior do domínio da função) - Digitar sobre o comando OK.

1.3) Ajustar os resultados para valores inteiros. Observar que todas as colunas juntas formam uma tabela.

ANEXO 02

A DISTRIBUIÇÃO NORMAL - ESTIMAÇÃO - INTERVALO DE CONFIANÇA - DIMENSIONAMENTO E ERRO.

2. A AMOSTRAGEM ALEATÓRIA SIMPLES E A DISTRIBUIÇÃO NORMAL

A distribuição de freqüências gerada pela repetição de um experimento, por diversas vezes, determina o que se chama função distribuição de probabilidades. Em princípio, a distribuição de freqüências é a descrição de uma população que é derivada da população origem. Ela descreve uma população na qual cada medida representa uma estatística calculada da distribuição origem. Existem diversos modelos que dependem do tipo de variável em estudo (discretas ou contínuas). As variáveis discretas são as relacionadas a atributos (qualidades), e as contínuas são variáveis relacionadas a medições (altura, diâmetro, concentração; área e outras). Diversos modelos de probabilidades poderiam ser exemplificados, mostra-se apenas a versatilidade da distribuição normal padrão (modelo contínuo), inclusive como aproximação de modelos discretos.

2.1 A DISTRIBUIÇÃO NORMAL

A distribuição normal foi descoberta, primeiramente, pelo matemático inglês DeMoivre (1667-1754). Mais tarde, foi redescoberta e aplicada às ciências naturais, bem como às sociais e, para assuntos práticos, pelo matemático francês Laplace (1749-1827). Foi também amplamente desenvolvida e utilizada pelo matemático, físico e astrônomo alemão Gauss (1777-1854). Um dos primeiros a fazer extenso uso da distribuição normal em estatística social foi o astrônomo e estatístico belga Quetelet (1796-1874). Um dos pioneiros na sua aplicação, com referência a dados biológicos, foi o antropologista, biometrista, criminologista, meteorologista, psicólogo e estatístico inglês Sir Francis Galton (1822-1911) -

primo de Charles Darwin. Galton expressou sua paixão pela distribuição normal, utilizando-se das seguintes palavras:

> "Não conheço nada tão apto a impressionar a imaginação como a maravilhosa forma da ordem cósmica definida pela "Lei da Freqüência de Erros". A lei teria sido personificada e endeusada pelos gregos se eles a tivessem conhecido. Ela impera serena e imperturbável no meio das maiores confusões. Quanto mais colossal a multidão e maior a anarquia aparente, mais perfeita é a sua oscilação. É a suprema lei do desarrazoado exorbitante. Sempre que uma grande amostra de elementos caóticos é tomada em consideração e disposta na ordem de suas magnitudes, surge uma bela e insuspeita forma de regularidade, prova que estivera latente todo o tempo[14]."

2.1.1 A Matemática da Distribuição Normal

Geralmente, as medidas (estatísticas) calculadas das amostras aleatórias tendem a ser distribuídas normalmente, mesmo que os dados originais não sejam normais. Apesar de existirem diversas exceções, qualquer estudo inicia-se da distribuição normal. A distribuição normal é uma das mais importantes variáveis aleatórias contínuas. Serve como uma excelente aproximação para uma grande classe de distribuições (discretas e contínuas), que têm grande importância na prática (Meyer, 1983).

Características da distribuição normal:

Definição: A variável aleatória X, que assume todos os valores reais $-\infty \leq x \leq \infty$, tem a distribuição normal (ou de Gauss) e sua função densidade de probabilidades é dada por:

$$f(x) = (1/ (\sigma\sqrt{2p}\ )\ )\ \exp\text{-}1/2\ [(x\text{-}\mu)/\sigma]^2\ ,\ -\infty \leq x \leq \infty \text{ e } \sigma>0 \qquad (2.1.1.1)$$

As notações utilizadas para se dizer que a f(x) representa uma distribuição normal são: $N(\mu,\sigma)$ e $N(\mu,\sigma^2)$. A média (μ) e o desvio-padrão (σ) populacionais representam os parâmetros da distribuição normal.

Propriedades da distribuição normal:

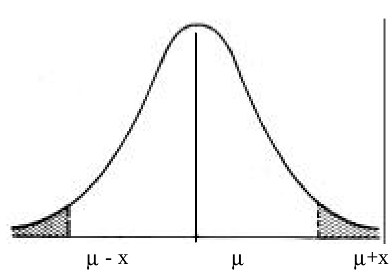
a) $E(x) = \mu$, a média da variável aleatória x é igual a μ.

b) $V(x) = \sigma^2$, a variância da variável aleatória x é igual a σ^2.

c) Forma de sino e simetria com relação ao eixo $x = \mu$ (veja figura 22).

d) $\int_{-\infty}^{+\infty} f(x)dx = 1$, isto quer dizer que a área abaixo da curva é igual a 1, e que a mesma representa a probabilidade total de um experimento.

FIGURA 22: Distribuição Normal.



---

[14] Citado por Helen M. Walker, Elementary Statistical Methods (Nova York: Henry Holt and Company, 1943).

Para cada valor de μ e σ, f(x) tem suas probabilidades particulares. Para contornar o inconveniente, criou-se uma v.a. Z, na qual $Z= (x-\mu)/\sigma$ e f(z) também é uma distribuição normal, mas com média (μ) = 0 e o desvio-padrão (σ) = 1. Por meio das propriedades da média e do desvio-padrão, verifica-se que, quando se subtrai uma constante de todos os valores, a média fica diminuída da própria constante, e que quando se dividem, por uma constante, todos os valores, o desvio-padrão fica dividido pela constante. Então, com a transformação $Z = (x-\mu)/\sigma$ tem-se a seguinte função:

$$f(z) = (1/\sqrt{2p}\ ) \exp-(1/2\ [z]^2)\ , \qquad -\infty \leq z \leq \infty \qquad (2.1.1.2)$$

Logo, f(z) é uma distribuição N(0 ; 1).

Em resumo, qualquer distribuição normal pode ser transformada em uma distribuição, também normal, com média zero, mudando-se para a direita ou para a esquerda uma parte da mesma, igual à sua média. Isso é realizado, fazendo-se (x_i - μ), que representa a subtração de μ de cada observação na população. Assim, as novas observações terão média zero, sendo que o desvio-padrão (σ), além de representar a dispersão da distribuição normal, pode ser utilizado como unidade de medida em cálculo de probabilidades.

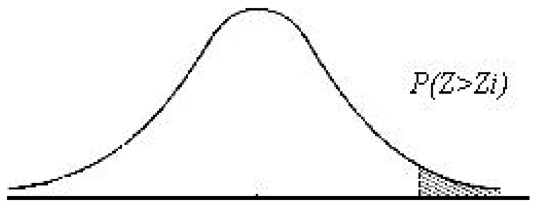
2.1.2 Cálculo de Probabilidades

Para o cálculo de probabilidades, é necessária a utilização de integrais, pois a probabilidade é representada pela área entre dois pontos. Com a utilização de programas computacionais (Matlab, Minitab, Excel e outros, veja anexo 03), fica mais fácil esse cálculo.

A tabela seguinte fornece as probabilidades de $z_i \leq Z \leq \infty$, isto é, $\int_{z_i}^{\infty} f(t)dt$. A função é a mesma f(z). Muda-se o símbolo em função do rigor matemático. A tabela 17 (gerada computacionalmente) mostra as probabilidades para os principais valores de Z.

Existem diversas formas para se representar a distribuição normal padronizada. Essa forma foi a escolhida devido à maior sintetização dos dados. Basta observar a probabilidade que está sendo procurada e ajustar o resultado ao da tabela.

TABELA 17: Probabilidades Para a Distribuição Normal - N(0;1)



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0666	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

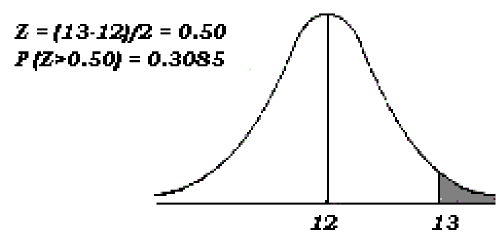
2.1.3 Um Exemplo de Cálculo de Probabilidades com o uso da Tabela.

Como exemplo hipotético, pode-se citar a seguinte distribuição, normalmente distribuída, relacionada com teor de alumínio, em determinada região, como sendo uma variável aleatória com distribuição N(12; 2)g/m^3 e as probabilidades procuradas como sendo: a) P(x >13g/m^3); b) P(x < 9g/m^3) ; c) P(11 g/m^3<x<13 g/m^3).

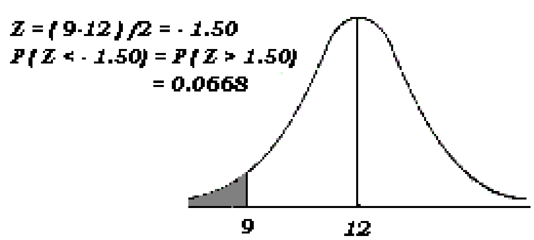
Primeiramente, observa-se que a distribuição do problema não é a distribuição normal padronizada (Z), portanto os valores devem ser transformados de tal forma que a média da distribuição se torne zero (0) e o desvio-padrão, um (1).

Resolução:

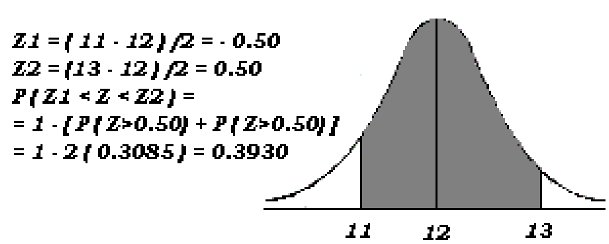
a) P(X >13g/m^3) = P (Z>0.50) = 0.3085



b) P(X < 9g/m^3) = 0.0668



c) P(11 g/m^3<X<13 g/m^3) = 0.3930



Observa-se que, para os cálculos das probabilidades, foram utilizadas duas propriedades da distribuição normal: a simetria em torno da média zero e a probabilidade total igual a 1.

2.2 O TEOREMA CENTRAL DO LIMITE

Se uma variável aleatória X puder ser representada pela soma de quaisquer n variáveis independentes, esta soma para "n" suficientemente grande terá distribuição aproximadamente normal. Esse importante resultado é conhecido como teorema central do limite. Esse teorema constitui uma clara generalização da aproximação de DeMoivre-Laplace, porque as variáveis aleatórias X_i, que assumiam apenas os valores 0 e 1, foram substituídas por variáveis aleatórias, que possuem qualquer espécie de distribuição (que tenham média e desvio-padrão finitos). O fato de que X_i possa ter (essencialmente) qualquer distribuição de probabilidades, e a soma $Y= X_1 + X_2 + ...+X_n$, possa ser aproximada por uma variável normalmente distribuída, exprime a importância fundamental da distribuição normal na teoria das probabilidades. Em muitos problemas, as "n" variáveis aleatórias independentes podem ser representadas pela soma (Y), então a distribuição de Y poderá ser aproximada pela distribuição normal (Meyer, 1984).

As condições gerais enunciadas no teorema central do limite podem ser colocadas resumidamente da seguinte forma: cada parcela da soma contribui com um valor sem importância para a variação da soma total, e é muito improvável que qualquer parcela isolada forneça uma contribuição muito grande para a soma total. Os erros de mensuração têm essa característica. O erro total pode ser representado como a soma de muitas contribuições pequenas, em que nenhuma das contribuições influi muito no erro total, e as parcelas não necessitam ser normalmente distribuídas para que a soma tenha uma distribuição normal (Meyer, 1983).

Sintetizando todas as informações, pode-se escrever o teorema do limite central da seguinte forma:

Teorema - Sejam X_1, X_2, ..., X_n, "n" variáveis aleatórias com mesma distribuição de probabilidades. Sejam $E(X_i) = \mu$ e $\sigma^2 = V(X_i)$ média e variância comuns. Se $Y = X_1 + X_2 + + \ldots + X_n$, então, $E(Y) = n\mu$ e $V(Y) = n\sigma^2$ (as propriedades para o valor esperado e para a variância foram utilizadas para a obtenção desses resultados), e, para "n" grande, a distribuição de Y será $N(n\mu; n\sigma^2)$. Ao se padronizar esta distribuição normal a v.a. Z assumirá a seguinte forma:

$$Z = (Y - n\mu) / (\sigma\sqrt{n}) \sim N(0;1) \tag{2.2.1}$$

O resultado acima é de suma importância para a verificação das distribuições dos parâmetros amostrais; a demonstração desse teorema está além do objetivo desse estudo.

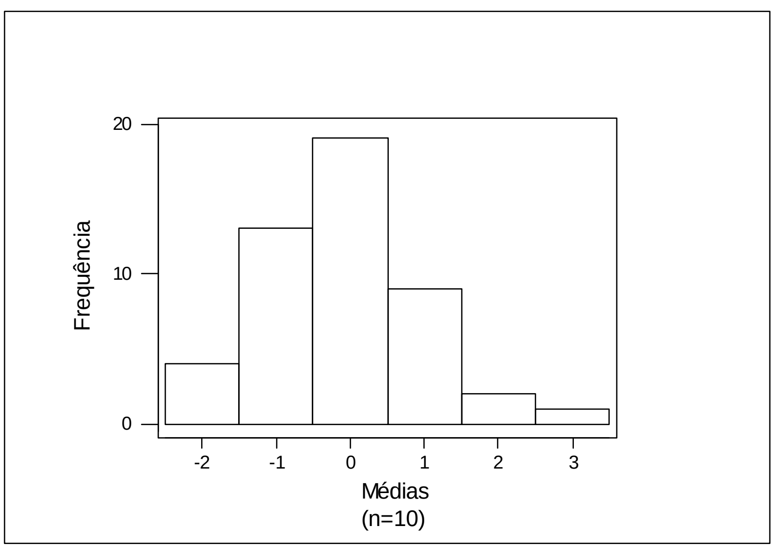
2.2.1 Um Exemplo de Aplicação do Teorema Central do Limite.

Para exemplificar, suponha-se que foram retiradas, de uma população uniforme, 50 amostras com 10 unidades em cada uma, e que o procedimento tenha sido repetido para amostras de tamanho 50 e 100 elementos. Os histogramas seguintes representam a distribuição das médias das amostras. Eles mostram que quanto maior o valor de n, mais a distribuição das médias tende à forma da distribuição normal (valores padronizados).

As principais estatísticas para cada grupo de médias são:

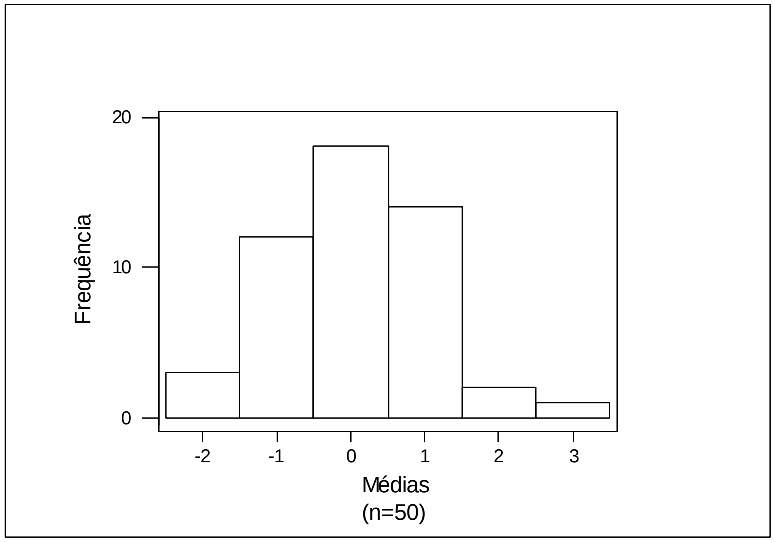
Variáveis	n	Média	Desvio-padrão
Amostras (10)	50	4.7940	1.0850
Amostras (50)	50	5.0262	0.3829
Amostras (100)	50	5.0940	0.3489

FIGURA 23: Distribuição das Médias para Amostras de 10 elementos



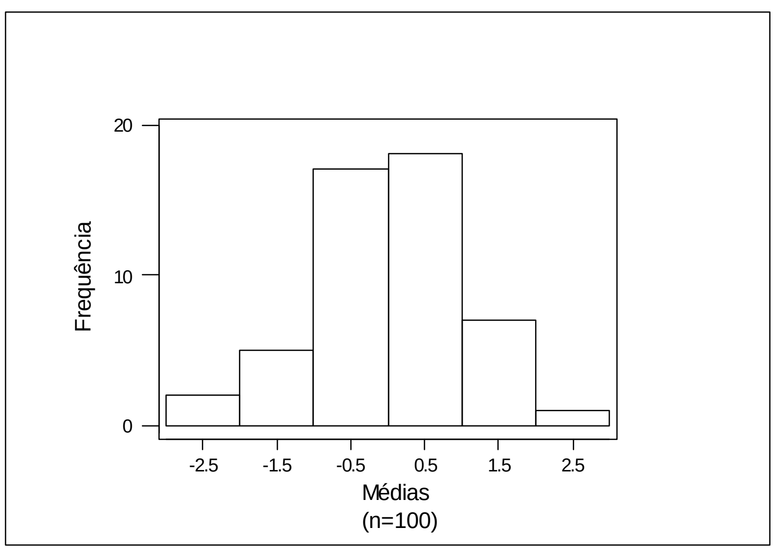
Fonte: EARLR / 1999.

FIGURA 24: Distribuição das Médias para Amostras de 50 elementos



Fonte: EARLR / 1999.

FIGURA 25: Distribuição das Médias para Amostras de 100 elementos



Fonte: EARLR / 1999.

Ao padronizar os valores pertencentes aos respectivos grupos de médias, as estatísticas são:

Variável	**n**	**Média**	**Desvio-padrão**
Amostras (10)	50	0.000	1.000
Amostras (50)	50	0.000	1.000
Amostras (100)	50	0.000	1.000

Observa-se que, além da forma normal da distribuição de frequências, as estatísticas confirmam os valores da população normal padrão. Normal com média zero e desvio-padrão igual a um, confirmando que, para uma população não normal, a média de amostras de tamanho "n" ($n \rightarrow \infty$) tende para uma normal. O domínio da distribuição uniforme original era de 0 a 10, implicando uma média igual a 5, e um desvio-padrão aproximado de 2.88.

2.2.2 Aproximação das Distribuições, das Variáveis Aleatórias de Interesse, pela Distribuição Normal.

Utilizando-se a distribuição normal, podem-se estudar tanto as variáveis qualitativas (atributos) quanto as variáveis quantitativas (medições), por meio das respectivas aproximações.

Nos levantamentos por amostragem, busca-se medir e registrar, para cada unidade amostral, determinadas características ou especificações. Uma seleção por amostragem pode ter diversas finalidades, mas, na maioria dos casos, o interesse concentra-se em quatro variáveis: média, valor total, relação entre dois valores médios ou totais, e proporção de unidades com determinada característica.

Nomenclatura e definições para parâmetros populacionais e estatísticas amostrais.

População	Amostra	Representação da estimativa
Média: $\mu_{y} = \sum_{i=1}^{N} y_i / N = \overline{Y}$	Média: $\overline{y} = \sum_{i=1}^{n} y_i / n$	$\overline{y}$
Total: $Y = \sum_{i=1}^{N} y_i = y_1 + y_2 + \ldots + y_N$	$\sum_{i=1}^{n} y_i = y_1 + y_2 + \ldots + y_n$	$N\overline{y}$
Índice da população: $R = \mu_y / \mu_x$	$\overline{y} / \overline{x} = \sum_{i=1}^{n} y_i / \sum_{i=1}^{n} xi$	$\overline{y} / \overline{x}$
Proporção: $P = X/N$	$\hat{p} = x/n$	$\hat{p}$

X = número de sucessos na população.

N = número de elementos da população.

y_i = valores da variável aleatória.

n = tamanho da amostra

O processo de cálculo da estimativa e o plano de amostragem utilizado influenciam diretamente na precisão da estimativa. Para qualquer fórmula utilizada, deve-se assegurar o processo de amostragem e o método de estimativa para os quais a fórmula foi estabelecida. Uma das características mais importantes de uma estimativa é a não tendenciosidade, isto é, o valor esperado da estatística amostral deve ser igual ao parâmetro populacional. Neste ponto, algumas perguntas devem ser respondidas: • A média da amostra é uma estimativa não tendenciosa da média populacional?. • E o desvio-padrão da amostra?. • A proporção da amostra pode ser tida como não tendenciosa?.

Se o método for sem tendência, esse resultado será verificado para qualquer população de valores finitos y_i, e para qualquer valor de n. As verificações de que as estatísticas amostrais são não tendenciosas serão feitas para o caso de uma amostra acidental (aleatória) simples.

Aqui, o estudo será direcionado para a variável "proporção de sucessos", que é uma medida qualitativa (resultante de uma contagem). As inferências estatísticas para as amostragens adaptadoras se encontram no anexo 07.

2.3 A DISTRIBUIÇÃO BINOMIAL

A distribuição Binomial representa acontecimentos com reposição, nos quais a probabilidade de sucesso "p" é constante para cada repetição do experimento. A variável aleatória x representa o número de sucessos para "n" repetições do experimento. Então, f(x) é dada por:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0\ ,\ 1\ ,\ 2\ ,\ \dots\dots\dots,\ n \qquad (2.3.1)$$

Outras características da distribuição Binomial são:

i) $E(x) = m = np$ (2.3.2)

ii) $V(x) = s^2(x) = np(1-p)$ (2.3.3)

iii) A distribuição para acontecimentos sem reposição (distribuição Hipergeométrica) pode ser aproximada pela Binomial, quando $N / n \leq 0.10$.

iv) Quando $n \to \infty$ e $np \geq 5$, a distribuição Binomial converge para a Normal.

Esta convergência é melhor visualizada quando $p \cong (1-p) \cong 0.5$, pois a distribuição Binomial é praticamente simétrica neste caso.

A proporção é um caso de média, no qual as observações assumem valores 0 ou 1. A média da distribuição de amostragem de proporções é a "população" média, ou a proporção populacional designada por "p". Se um número suficientemente grande de amostras é extraído, a média das proporções das amostras $\hat{p}$ tende para p. Desta forma, $\hat{p}$ é não tendencioso. A distribuição Binomial é outra maneira de se verificar a distribuição de uma proporção. Para "n" grande, a distribuição Binomial converge[15] para a distribuição normal em que:

$$Z = ((x \pm 0.5) - \mu)/\sigma \sim N(0\,;\,1) \qquad (2.3.4)$$

O valor 0.5 acima representa a correção de continuidade, pois a distribuição Binomial é discreta, e a estamos aproximando pela Normal, que é uma função contínua. Dividindo ambos os termos por n ($n \to \infty$), chega-se a:

$$Z = ((x/n \pm 0.5/n) - \mu/n)/(\sigma/n) \qquad (2.3.5)$$

Uma v.a. distribuída binomialmente tem média μ = np e desvio-padrão σ = npq. Substituindo-se estes resultados na fórmula acima, tem-se que:

$$Z = ((x/n \pm 0.5/n) - np/n)/(\sqrt{npq/n^2}) \text{ então,}$$

[15] Veja "Triola, 1999; p.126".

$$Z = (\hat{p} - p)/\sqrt{pq/n} \sim N(0\,;1). \tag{2.3.6}$$

O desvio-padrão[16] $\sqrt{pq/n}$ pode ser corrigido conforme a cpf (correção das populações finitas). Para isto, basta multiplicá-lo pela raiz de (N-n)/(n-1).

Demonstração:

Seja uma a.a de "n" elementos retirada de uma população B(n ; p). Seja $\hat{p}$ = x/n (proporção de sucessos da amostra), então,

$$E(\hat{p}) = E(x/n) = np/n = p \tag{2.3.7}$$

Logo, $\hat{p}$ é um estimador não tendencioso de p, e a variância de p será dada por:

$$V(\hat{p}) = V(x/n) = (1/n^2)\,npq = pq/n. \tag{2.3.8}$$

Como $\hat{p}$ é um estimador não tendencioso de p, a variância estimada poderá ser escrita da seguinte forma:

$$S^2(\hat{p}) = \hat{p}\hat{q}/n \text{ (variância da amostra).} \tag{2.3.9}$$

Assim, a variável aleatória $\hat{p}$ tem distribuição N(p; $\sqrt{pq/n}$) e, padronizando, chega-se ao seguinte resultado:

$$Z = (\hat{p} - p)/\sqrt{pq/n} \tag{2.3.10}$$

A condição de que np ≥ 5 é suficiente para a aproximação normal na maioria dos objetivos práticos (Wallis & Roberts, 1964).

[16] Quanto mais próximo de 0.5 estiver o valor de p, maior será o desvio-padrão (veja figura 26).

Agora, existem condições teóricas para se estimar a proporção populacional, como também para se fazer um estudo sobre dimensionamento.

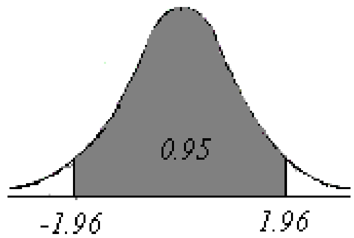
2.3.1 Um Intervalo de Confiança Para a Proporção Populacional "p".

"Seja $\hat{q}$ um estimador não tendencioso de q distribuído normalmente, isto é, $\hat{q}$ tem distribuição N(q ; $s^2(\hat{q})$), então a v.a. $Z = \dfrac{\hat{q} - q}{\sqrt{s^2(\hat{q})}}$ tem distribuição normal padrão, e um intervalo de confiança de $100(1-a)\%$ para q será dado por:

$$[\hat{q} - Z\sqrt{s^2(\hat{q})}\ ; \hat{q} + Z\sqrt{s^2(\hat{q})}\ ] \qquad (2.3.1.1)$$

Demonstração: A v.a. Z tem distribuição normal padrão, isto é, N(0;1). Então, a $P(z_0 \leq Z \leq z_1) = 1 - a$.

Nota-se que z_0 e z_1 são valores da distribuição normal (iguais a menos do sinal e simétricos). Por exemplo, se o nível de confiabilidade é de 95%, os valores de z_0 e z_1 (que entre ambos existe uma probabilidade de 0.95) são respectivamente -1.96 e +1.96. Assim, acima de z_1, tem-se uma probabilidade de 0.025 $(a/2)$, que é a mesma probabilidade abaixo de z_0. Logo, $a = 0.05$.



Portanto, ao substituir a v.a. Z na probabilidade, tem-se que:

$$P(z_0 \leq \frac{\hat{q} - q}{\sqrt{s^2(\hat{q})}} \leq z_1) = 1 - a \qquad (2.3.1.2)$$

Explicitando q (parâmetro populacional a ser estimado) no centro do intervalo, chega-se ao seguinte resultado:

$$P(\hat{q} - z_1 \sqrt{s^2(\hat{q})} \leq q \leq \hat{q} + z_0 \sqrt{s^2(\hat{q})}) = 1 - a \qquad (2.3.1.3)$$

Assim, para qualquer estatística amostral com distribuição normal, pode-se estimar o parâmetro populacional através de um intervalo de $100(1-a)\%$ de confiabilidade" (Thompson,1992)[17] .

A literatura atual representa o intervalo de confiança da seguinte forma:

$$\hat{q} \pm z_1 \sqrt{s^2(\hat{q})} \qquad (2.3.1.4)$$

Portanto, um intervalo de 100(1-α)% para a proporção populacional será dado por:

---

[17] Tradução Livre.

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\hat{q}}{n}} \qquad (2.3.1.5)$$

Nota-se que a cpf foi omitida na construção do intervalo, e que o desvio-padrão amostral foi utilizado como estimativa do desvio-padrão populacional.

2.3.2 Dimensionamento - O Caso da Proporção Populacional.

"Ao se fazer um planejamento, existe a preocupação com o tamanho da amostra, o que nem sempre é tão simples como o desejado.

Supondo que se deseje estimar o parâmetro populacional q através de um estimador $\hat{q}$, de modo que a probabilidade deste estar próximo do valor verdadeiro seja elevada. Especificar uma diferença máxima (d) entre a estimativa e o valor verdadeiro é permitir uma pequena probabilidade (α) de que o erro possa exceder esta diferença. O objetivo é escolher um tamanho para a amostra (n) tal que:

$$P(|\hat{q} - q| > d) \leq \alpha. \qquad (2.3.2.1)$$

Se $\hat{q}$ for um estimador não tendencioso de q e tiver distribuição normal, em que $Z = \dfrac{\hat{q} - q}{\sqrt{s^2(\hat{q})}}$ e associando-se a probabilidade com o ponto superior da distribuição normal padrão, tem-se que:

$$P\left(\left|\frac{\hat{q} - q}{\sqrt{s^2(\hat{q})}}\right| > z\right) = P\left(|\hat{q} - q| > z\sqrt{s^2(\hat{q})}\right) = \alpha. \qquad (2.3.2.2)$$

A variância do estimador $\hat{q}$ diminui com o aumento de "n", de modo que a desigualdade acima será satisfatória se n for suficientemente grande para tornar z $\sqrt{S^2(\hat{q})} \leq$ d. Como exemplo, o tamanho da amostra (n) para se estimar a média populacional, através de uma a.a. simples, é desenvolvido da seguinte forma:

Z $\sqrt{(\frac{(N-n)}{N}\frac{S^2}{n}} = d$, então, a dimensão da amostra será dada por:

$$n = \frac{1}{(\frac{d^2}{Z^2S^2} + \frac{1}{N})} = \frac{1}{\frac{1}{n_0} + \frac{1}{N}} \quad \text{onde } n_0 = \frac{Z^2S^2}{d^2} \; . \tag{2.3.2.3}$$

Se a cpf não for importante, a fórmula acima se reduz para

$n = (\frac{ZS}{d})^2$ " (Cochran, 1977)[18]. (2.3.2.4)

Para um estudo, no qual o interesse está na proporção de sucessos, tem-se que:

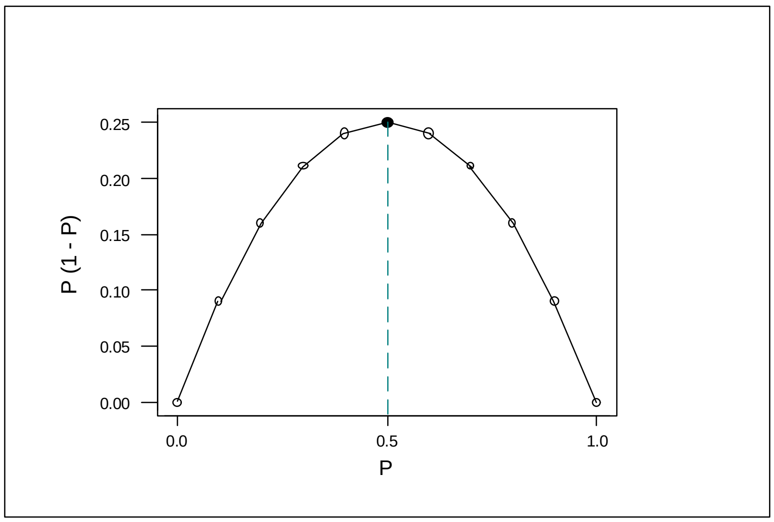
$$n = (\frac{Z}{d})^2 \, \hat{p}(1-\hat{p}) \; , \quad \text{visto que } S = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \; . \tag{2.3.2.5}$$

Observa-se que a proporção da amostra nem sempre é conhecida. Desta forma, a solução estaria em superestimar "n".

[18] Tradução Livre.

Sabe-se que $p(1-p) = p - p^2$, e que $0 \leq p \leq 1$, desta forma pode-se fazer um estudo para a proporção populacional "p" (veja figura 26). Observa-se que $p - p^2$ representa uma função quadrática.

FIGURA 26: Dimensionamento Utilizando-se o Ponto de Máximo de $(p - p^2)$**.**

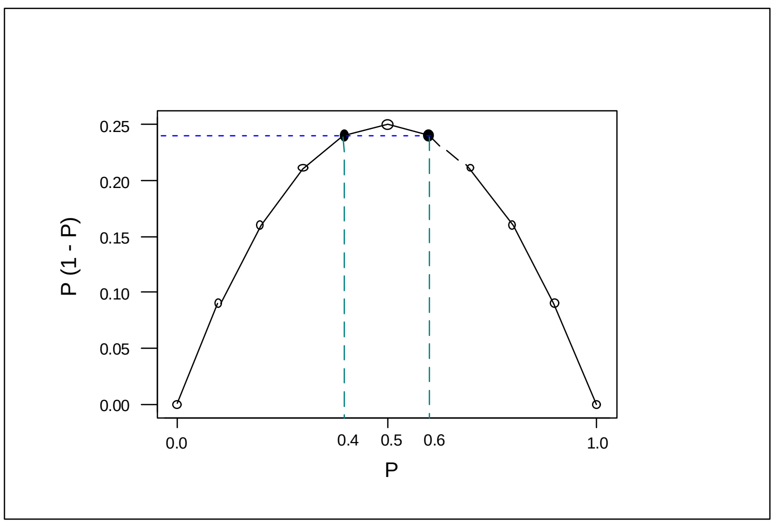


O ponto de máximo de $f(p) = p - p^2$ ocorre quando $p = 0{,}5$ (veja figura 24), desta forma, $n = (\frac{Z}{d})^2 \frac{1}{2}\frac{1}{2} = (\frac{Z}{2d})^2$. Assim, para um erro (d) de 0,05 e um nível de confiabilidade de 95% , $n = (\frac{1{,}96}{0{,}10})^2 = 19{,}6^2 \cong 385$ elementos. Se fosse necessária a utilização da c.p.f, e a população total fosse de 20.000 elementos, $n_f = \frac{N.n}{N+n} \cong 378$ elementos (n_f = tamanho final da amostra).

Para se diminuir o valor de "n" , ou se diminui o nível de confiabilidade, ou se aumenta o erro aceitável em torno do valor central. Esta afirmação é válida, independentemente de se ter uma estimativa para "p". No anexo 04 encontram-se tabelas com os valores de "n" para diversos valores de "p".

Quando não se conhece uma estimativa para a proporção populacional "p", o valor de "n" fica superestimado, e isto se deve à utilização do ponto de máximo da parábola, conforme visto anteriormente. Quando se distancia de p para a esquerda (valor zero) ou direita (valor 1), o valor de "n" diminui, portanto, é interessante o conhecimento de um valor $\hat{p}$ (estimativa de p). O tamanho da amostra será o mesmo para valores simétricos em torno do valor 0,5 (veja figura 27).

FIGURA 27: Valores de "n" Para Valores Simétricos em Torno de P = 0,5 .



Fonte: EARLR / 2000

Pode-se observar que para $p = 0{,}4$ ou $p = 0{,}6$, o resultado de $p(1-p)$ será de 0,24 para qualquer um dos dois valores. Assim, se o nível de confiabilidade for de 95% e o erro de 0,05 , o tamanho da amostra será dado por: $n = (\frac{Z}{d})\hat{p}(1-\hat{p}) = \; = (\frac{1.96}{0{,}05})^2.0{,}4.0{,}6 \cong 369$

elementos. Utilizando-se a c.p.f, para uma população de N=20.000, o tamanho da amostra será de $n_f \cong 363$ elementos.

2.4 DIMENSIONAMENTO E ERRO.

Do tópico anterior, sabe-se que $d = Z\sqrt{\frac{pq}{n}}$, daí vem a pergunta: Como se comporta o erro com relação à "n"? Supondo-se um nível de confiabilidade fixo, por exemplo um nível de 95% de confiabilidade, em que Z=1.96, pode-se construir a tabela 18, variando-se n e p, para se fazer o cálculo de "d".

Observa-se que o erro diminui quando se aumenta o valor de "n" e que quando $n \to \infty$, o erro tende a ser constante. A figura 28 representa melhor esta associação de "n", "d" e p. Outro fato a ser observado é que quando se aumenta o nível de confiabilidade, o valor de Z aumenta, levando a um aumento do erro (maior confiabilidade e menor precisão).

A correção de continuidade deve ser feita apenas para pequenas populações, pois a diferença é nula quando $N \to \infty$.

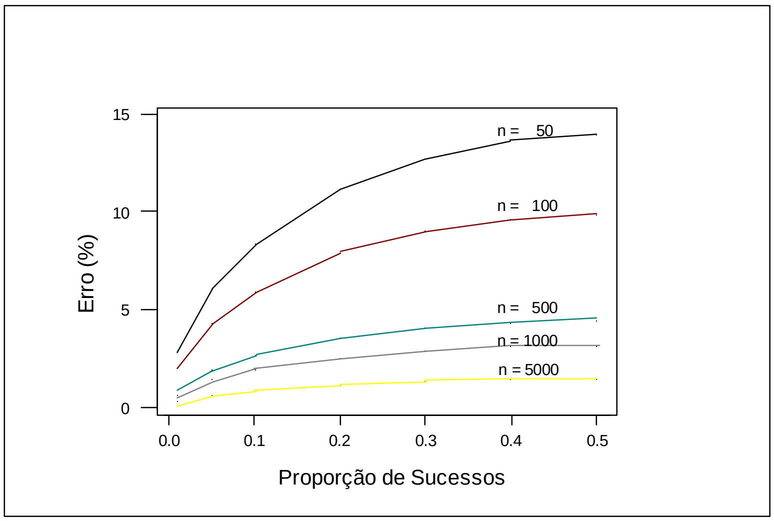
Todo o estudo foi feito, supondo-se a retirada de uma amostra aleatória simples. Existem outros tipos de amostragem que resultam num menor desvio-padrão e, conseqüentemente, em um menor erro (d). Outro fato a ser verificado é o de que a variância pode ter distribuição não normal, e isto pode afetar o resultado final.

TABELA 18: Erro (%) para Alguns Valores de "p" e "n" - Nível de 95% de Confiabilidade.

Valores de p	Valores de n				
	50	**100**	**500**	**1000**	**5000**
0.01	2.76	1.95	0.87	0.60	0.27
0.05	6.04	4.27	1.91	1.35	0.60
0.10	8.32	5.88	2.63	1.86	0.83
0.20	11.09	7.84	3.51	2.48	1.11
0.30	12.70	8.98	4.02	2.84	1.27
0.40	13.58	9.60	4.29	3.04	1.36
0.50	13.86	9.80	4.38	3.09	1.39

Fonte: E.A.R.L.R / 2000.

FIGURA 28: Tamanho da Amostra e Erro (%).



Fonte: E.A.R.L.R / 2000.

Os outros tipos de amostragem, como as adaptadoras, são exemplos de como se obter menor desvio-padrão (veja anexo 07).

Para o cálculo das estimativas para as seleções clássicas de amostragem, já existe no mercado o software "Sampling", que foi desenvolvido pelo Departamento de Estatística da UFMG, com o apoio da FAPEMIG. Esse software funciona como um complemento do software estatístico "Minitab" e permite ao usuário a estimação de parâmetros populacionais, por meio de métodos estatísticos de estimação pontual e intervalar.

O "Sampling" oferece alguns recursos, como:

- a determinação de tamanho para a amostra (n), considerando-se a minimização do custo final e da variabilidade dos estimadores;
- a análise descritiva e gráfica global e/ou por estrato;
- a estimação dos parâmetros populacionais por meio dos métodos usuais (razão e regressão).

ANEXO 03

DISTRIBUIÇÕES DE PROBABILIDADES - NORMAL E STUDENT

O objetivo deste anexo é mostrar a utilização dos "softwares" Maple e Minitab na montagem de tabelas da distribuição Normal e Student, respectivamente.

3.1 Cálculo das Probabilidades Normais utilizando o "Software" Maple

$$P(Z \geq z_i) = \int_{zi}^{\infty} f(t)dt \;=\; \int_{zi}^{\infty} \left(1/\sqrt{2p}\right) e^{-(1/2)t^2}\, dt$$

O "software" maple é de fácil manuseio. O leitor precisa apenas conhecer os comandos básicos. Primeiramente, entra-se com a fórmula matemática e com os limites de integração.

```
 Int(exp(-z^2/2)/(sqrt(2*pi)),z=0..0.01);
➢
```

```
        .01
       /                  2   1/2
      |        exp(- 1/2 z ) 2
      |    1/2 ------------------ dz
      |               1/2
     /              pi
       0
```

O segundo passo é colocar a letra p de pi em maiúscula, para que a integral da função não forneça o valor de *p* como se fosse uma incógnita. Assim, a integral da distribuição normal padrão entre (0.00 e 0.01) será:

```
> int(exp(-z^2/2)/(sqrt(2*Pi)),z=0..0.01);
```

```
                    .003989356314
```

Essa integral representa a probabilidade de Z estar entre (0.00 e 0.01), portanto, ao se subtrair 0.5000, o valor encontrado será igual à probabilidade de Z maior ou igual a 0.01.

```
> 0.5-0.003989356314;
```

.4960106437

O valor encontrado tem 10 casas decimais e a tabela foi gerada com 4 casas decimais. Portanto, fixa-se o número de casas decimais.

> evalf(",4);

.4960

O valor 0.4960 é o valor da probabilidade procurada e, para se montar a tabela, basta ir mudando os valores do limite superior da integral, isto é, de 0.01 para 0.02, de 0.02 para 0.03,........, até o valor 3.00. O valor 3.00 foi o escolhido, por ser o valor máximo utilizado para Z em inferência estatística (99.7%). Veja tabela 01 do tópico 3.

3.2 Cálculo das Probabilidades da Distribuição de Student - "Software" Minitab

$$P(T \geq t_i) = \int_{ti}^{\infty} ((1/\sqrt{kp})\Gamma((k+1)/2))/((\Gamma(k/2)((1+l^2/k)^{(k+1)/2}\partial l$$

Para gerar as probabilidades pelo "software" Minitab, é necessário observar os seguintes passos:

3.2.1) Colocar na coluna 1 (C1) os valores da estatística T.

3.2.2) Com o mouse, clicar no menu principal o comando Calc e os seguintes comandos: Probability - Distributions - T - Cumulative Probability - Degrees of freedom (digita-se no quadrado o n$^{\underline{o}}$ de graus de liberdade) - Input Column (digite no quadrado C1) - Optional Storage (digita-se no quadrado a coluna onde os resultados devem aparecer) - Clicar sobre o comando OK. Os resultados das probabilidades cumulativas aparecerão na planilha na coluna especificada. Para obter o resultado da integral acima, basta fazer o cálculo (1 - o valor da probabilidade encontrada).Veja tabela 19.

TABELA 19 - Distribuição de Student - "Gosset"

g.l	0.250	0.100	0.050	0.025	0.010	0.005
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.817	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.307
7	0.711	1.415	1.895	2.365	2.998	3.500
8	0.706	1.397	1.860	2.306	2.897	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.813	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.696	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.625	2.977
15	0.691	1.341	1.753	2.132	2.601	2.947
16	0.690	1.337	1.746	2.120	2.584	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.540	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.320	1.714	2.067	2.500	2.807
24	0.684	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
31	0.692	1.310	1.696	2.040	2.453	2.744
32	0.682	1.309	1.694	2.037	2.449	2.739
33	0.682	1.309	1.692	2.035	2.445	2.733
34	0.682	1.308	1.691	2.032	2.441	2.728
35	0.682	1.307	1.690	2.030	2.438	2.724
36	0.681	1.306	1.688	2.028	2.435	2.720
37	0.681	1.306	1.687	2.026	2.431	2.715
38	0.681	1.305	1.686	2.024	2.429	2.712
39	0.681	1.304	1.685	2.023	2.426	2.708
40	0.681	1.304	1.684	2.021	2.423	2.705
41	0.681	1.303	1.683	2.020	2.421	2.701
42	0.680	1.302	1.682	2.018	2.419	2.698
43	0.680	1.302	1.681	2.017	2.416	2.695
44	0.680	1.301	1.680	2.015	2.414	2.692
45	0.680	1.301	1.679	2.014	2.412	2.690
46	0.680	1.300	1.679	2.013	2.410	2.687
47	0.680	1.300	1.678	2.012	2.408	2.685
48	0.680	1.299	1.677	2.011	2.407	2.682
49	0.680	1.299	1.677	2.010	2.405	2.680
50	0.679	1.299	1.676	2.009	2.403	2.678
51	0.679	1.298	1.675	2.008	2.402	2.676
52	0.679	1.298	1.675	2.007	2.400	2.673
53	0.679	1.298	1.674	2.006	2.399	2.672
54	0.679	1.297	1.674	2.005	2.397	2.670
55	0.679	1.297	1.673	2.004	2.396	2.668
56	0.679	1.297	1.673	2.003	2.395	2.667
57	0.679	1.297	1.672	2.003	2.394	2.665
58	0.679	1.296	1.672	2.002	2.392	2.663
59	0.679	1.296	1.671	2.001	2.391	2.662
60	0.679	1.296	1.671	2.000	2.390	2.660

ANEXO 04

DIMENSIONAMENTO - NÍVEL DE 95% DE CONFIABILIDADE E ERRO DE 0.05 - Com e sem correção das populações finitas (CPF)

TABELA 20: Número de Elementos Para a Amostra - p=0.05 a p=0.25

N	0.05-sc	0.05-cc	0.10-sc	0.10-cc	0.15-sc	0.15-cc	0.20-sc	0.20-cc	0.25-sc	0.25-cc
1000	73	68	138	121	196	164	246	197	288	224
1100	73	68	138	123	196	166	246	201	288	228
1200	73	69	138	124	196	168	246	204	288	232
1300	73	69	138	125	196	170	246	207	288	236
1400	73	69	138	126	196	172	246	209	288	239
1500	73	70	138	126	196	173	246	211	288	242
1600	73	70	138	127	196	175	246	213	288	244
1700	73	70	138	128	196	176	246	215	288	246
1800	73	70	138	128	196	177	246	216	288	248
1900	73	70	138	129	196	178	246	218	288	251
2000	73	70	138	129	196	179	246	219	288	252
2100	73	71	138	129	196	179	246	220	288	253
2200	73	71	138	130	196	180	246	221	288	255
2300	73	71	138	130	196	181	246	222	288	256
2400	73	71	138	130	196	181	246	223	288	257
2500	73	71	138	131	196	182	246	224	288	258
2600	73	71	138	131	196	182	246	225	288	259
2700	73	71	138	131	196	183	246	225	288	260
2800	73	71	138	132	196	183	246	226	288	261
2900	73	71	138	132	196	184	246	227	288	262
3000	73	71	138	132	196	184	246	227	288	263
3100	73	71	138	132	196	184	246	228	288	264
3200	73	71	138	132	196	185	246	228	288	264
3300	73	71	138	132	196	185	246	229	288	265
3400	73	71	138	133	196	185	246	229	288	266
3500	73	72	138	133	196	186	246	230	288	266
3600	73	72	138	133	196	186	246	230	288	267
3700	73	72	138	133	196	186	246	231	288	267
3800	73	72	138	133	196	186	246	231	288	268
3900	73	72	138	133	196	187	246	231	288	268
4000	73	72	138	133	196	187	246	232	288	269
4100	73	72	138	134	196	187	246	232	288	269
4200	73	72	138	134	196	187	246	232	288	270
4300	73	72	138	134	196	187	246	233	288	270
4400	73	72	138	134	196	188	246	233	288	270
4500	73	72	138	134	196	188	246	233	288	271
4600	73	72	138	134	196	188	246	234	288	271
4700	73	72	138	134	196	188	246	134	288	271
4800	73	72	138	134	196	188	246	234	288	272
4900	73	72	138	134	196	189	246	234	288	273
5000	73	72	138	134	196	189	246	234	288	273
5100	73	72	138	134	196	189	246	235	288	273
5200	73	72	138	134	196	189	246	235	288	273
5300	73	72	138	134	196	190	246	235	288	274
5400	73	72	138	135	196	190	246	235	288	274
5500	73	72	138	135	196	190	246	235	288	274
5600	73	72	138	135	196	190	246	236	288	274
5700	73	72	138	135	196	190	246	236	288	275
5800	73	72	138	135	196	190	246	236	288	275
5900	73	72	138	135	196	190	246	236	288	275
6000	73	72	138	135	196	190	246	236	288	275
6100	73	72	138	135	196	190	246	236	288	276
6200	73	72	138	135	196	190	246	237	288	276
6300	73	72	138	135	196	190	246	237	288	276
6400	73	72	138	135	196	190	246	237	288	276
6500	73	72	138	135	196	190	246	237	288	276
6600	73	72	138	135	196	190	246	237	288	276
6700	73	72	138	135	196	190	246	237	288	276
6800	73	72	138	135	196	191	246	237	288	276
6900	73	72	138	135	196	191	246	238	288	276
7000	73	72	138	135	196	191	246	238	288	277
7100	73	72	138	135	196	191	246	238	288	277
7200	73	72	138	135	196	191	246	238	288	277
7300	73	72	138	135	196	191	246	238	288	278
7400	73	72	138	135	196	191	246	238	288	278
7500	73	72	138	135	196	191	246	238	288	278

N	Continuação da Tabela 20 0.05-sc	0.05-cc	0.10-sc	0.10-cc	0.15-sc	0.15-cc	0.20-sc	0.20-cc	0.25-sc	0.25-cc
7600	73	72	138	136	196	191	246	238	288	278
7700	73	72	138	136	196	191	246	238	288	278
7800	73	72	138	136	196	191	246	238	288	278
7900	73	72	138	136	196	191	246	239	288	278
8000	73	72	138	136	196	191	246	239	288	278
8100	73	72	138	136	196	191	246	239	288	279
8200	73	72	138	136	196	191	246	239	288	279
8300	73	72	138	136	196	191	246	239	288	279
8400	73	72	138	136	196	192	246	239	288	279
8500	73	72	138	136	196	192	246	239	288	279
8600	73	72	138	136	196	192	246	239	288	279
8700	73	72	138	136	196	192	246	239	288	279
8800	73	72	138	136	196	192	246	239	288	279
8900	73	72	138	136	196	192	246	239	288	279
9000	73	72	138	136	196	192	246	239	288	280
9100	73	72	138	136	196	192	246	240	288	280
9200	73	72	138	136	196	192	246	240	288	280
9300	73	72	138	136	196	192	246	240	288	280
9400	73	72	138	136	196	192	246	240	288	280
9500	73	72	138	136	196	192	246	240	288	280
9600	73	72	138	136	196	192	246	240	288	280
9700	73	72	138	136	196	192	246	240	288	280
9800	73	72	138	136	196	192	246	240	288	280
9900	73	72	138	136	196	192	246	240	288	280
10000	73	72	138	136	196	192	246	240	288	280
10100	73	72	138	136	196	192	246	240	288	280
10200	73	72	138	136	196	192	246	240	288	281
10300	73	72	138	136	196	192	246	240	288	281
10400	73	72	138	136	196	192	246	240	288	281
10500	73	72	138	136	196	192	246	240	288	281
10600	73	73	138	136	196	192	246	240	288	281
10700	73	73	138	136	196	192	246	240	288	281
10800	73	73	138	136	196	193	246	241	288	281
10900	73	73	138	136	196	193	246	241	288	281
11000	73	73	138	136	196	193	246	241	288	281
12000	73	73	138	136	196	193	246	241	288	282
13000	73	73	138	137	196	193	246	241	288	282
14000	73	73	138	137	196	193	246	242	288	283
15000	73	73	138	137	196	193	246	242	288	283
16000	73	73	138	137	196	194	246	242	288	283
17000	73	73	138	137	196	194	246	242	288	284
18000	73	73	138	137	196	194	246	243	288	284
19000	73	73	138	137	196	194	246	243	288	284
20000	73	73	138	137	196	194	246	243	288	284
21000	73	73	138	137	196	194	246	243	288	285
22000	73	73	138	137	196	194	246	243	288	285
23000	73	73	138	137	196	194	246	243	288	285
24000	73	73	138	137	196	194	246	244	288	285
25000	73	73	138	137	196	194	246	244	288	285
26000	73	73	138	137	196	195	246	244	288	285
27000	73	73	138	137	196	195	246	244	288	285
28000	73	73	138	137	196	195	246	244	288	286
29000	73	73	138	137	196	195	246	244	288	286
30000	73	73	138	137	196	195	246	244	288	286
40000	73	73	138	138	196	195	246	244	288	286
50000	73	73	138	138	196	195	246	244	288	286
60000	73	73	138	138	196	195	246	245	288	287
70000	73	73	138	138	196	195	246	245	288	287
80000	73	73	138	138	196	196	246	245	288	287
90000	73	73	138	138	196	196	246	245	288	288
100000	73	73	138	138	196	196	246	245	288	288
150000	73	73	138	138	196	196	246	246	288	288
200000	73	73	138	138	196	196	246	246	288	288
250000	73	73	138	138	196	196	246	246	288	288
300000	73	73	138	138	196	196	246	246	288	288

Fonte: EARLR / 1999

TABELA 21 : Número de Elementos Para a Amostra - p=0.30 a p=0.50

N	0.30-sc	0.30-cc	0.35-sc	0.35-cc	0.40-sc	0.40-cc	0.45-sc	0.45-cc	0.50-sc	0.50-cc
1000	323	245	350	260	369	270	381	276	384	278
1100	323	250	350	266	369	277	381	283	384	285
1200	323	255	350	271	369	283	381	290	384	291
1300	323	259	350	276	369	288	381	295	384	297
1400	323	263	350	280	369	293	381	300	384	302
1500	323	266	350	284	369	297	381	304	384	306
1600	323	269	350	288	369	300	381	308	384	310
1700	323	272	350	291	369	304	381	312	384	314
1800	323	274	350	294	369	307	381	315	384	317
1900	323	277	350	296	369	309	381	318	384	320
2000	323	279	350	298	369	312	381	321	384	323
2100	323	280	350	300	369	314	381	323	384	325
2200	323	282	350	302	369	316	381	325	384	327
2300	323	284	350	304	369	318	381	327	384	330
2400	323	285	350	306	369	320	381	329	384	332
2500	323	287	350	308	369	322	381	331	384	333
2600	323	288	350	309	369	324	381	333	384	335
2700	323	289	350	310	369	325	381	334	384	337
2800	323	290	350	312	369	327	381	336	384	338
2900	323	291	350	313	369	328	381	337	384	340
3000	323	292	350	314	369	329	381	339	384	341
3100	323	293	350	315	369	330	381	340	384	342
3200	323	294	350	316	369	331	381	341	384	343
3300	323	295	350	317	369	332	381	342	384	344
3400	323	295	350	318	369	333	381	343	384	346
3500	323	296	350	319	369	334	381	344	384	347
3600	323	297	350	319	369	335	381	345	384	347
3700	323	298	350	320	369	336	381	346	384	348
3800	323	298	350	321	369	337	381	347	384	349
3900	323	299	350	322	369	338	381	348	384	350
4000	323	299	350	322	369	338	381	348	384	351
4100	323	300	350	323	369	339	381	349	384	352
4200	323	300	350	324	369	340	381	350	384	352
4300	323	301	350	324	369	340	381	350	384	353
4400	323	301	350	325	369	341	381	351	384	354
4500	323	302	350	325	369	342	381	352	384	354
4600	323	302	350	326	369	342	381	352	384	355
4700	323	303	350	326	369	343	381	353	384	355
4800	323	303	350	327	369	343	381	353	384	356
4900	323	304	350	327	369	344	381	354	384	357
5000	323	304	350	328	369	344	381	355	384	357
5100	323	304	350	328	369	345	381	355	384	358
5200	323	305	350	328	369	345	381	355	384	358
5300	323	305	350	329	369	345	381	356	384	359
5400	323	305	350	329	369	346	381	356	384	359
5500	323	306	350	330	369	346	381	357	384	359
5600	323	306	350	330	369	347	381	357	384	360
5700	323	306	350	330	369	347	381	358	384	360
5800	323	306	350	331	369	347	381	358	384	360
5900	323	307	350	331	369	348	381	358	384	360
6000	323	307	350	331	369	348	381	359	384	361
6100	323	307	350	331	369	348	381	359	384	362
6200	323	308	350	332	369	349	381	359	384	362
6300	323	308	350	332	369	349	381	360	384	362
6400	323	308	350	332	369	349	381	360	384	363
6500	323	308	350	333	369	350	381	360	384	363
6600	323	308	350	333	369	350	381	361	384	363
6700	323	309	350	333	369	350	381	361	384	364
6800	323	309	350	333	369	351	381	361	384	364
6900	323	309	350	334	369	351	381	362	384	364
7000	323	309	350	334	369	351	381	362	384	365
7100	323	309	350	334	369	351	381	362	384	365
7200	323	310	350	334	369	352	381	362	384	365
7300	323	310	350	334	369	352	381	363	384	365
7400	323	310	350	335	369	352	381	363	384	366

N	0.30-sc	0.30-cc	0.35-sc	0.35-cc	0.40-sc	0.40-cc	0.45-sc	0.45-cc	0.50-sc	0.50-cc
	Continuação da tabela 21									
7500	323	310	350	335	369	352	381	363	384	366
7600	323	310	350	335	369	352	381	363	384	366
7700	323	310	350	335	369	353	381	364	384	366
7800	323	311	350	335	369	353	381	364	384	366
7900	323	311	350	336	369	353	381	364	384	367
8000	323	311	350	336	369	353	381	364	384	367
8100	323	311	350	336	369	353	381	364	384	367
8200	323	311	350	336	369	354	381	365	384	367
8300	323	311	350	336	369	354	381	365	384	368
8400	323	312	350	336	369	354	381	365	384	368
8500	323	312	350	337	369	354	381	365	384	368
8600	323	312	350	337	369	354	381	365	384	368
8700	323	312	350	337	369	354	381	366	384	368
8800	323	312	350	337	369	355	381	366	384	368
8900	323	312	350	337	369	355	381	366	384	369
9000	323	312	350	337	369	355	381	366	384	369
9100	323	312	350	338	369	355	381	366	384	369
9200	323	313	350	338	369	355	381	366	384	369
9300	323	313	350	338	369	355	381	367	384	369
9400	323	313	350	338	369	356	381	367	384	369
9500	323	313	350	338	369	356	381	367	384	370
9600	323	313	350	338	369	356	381	367	384	370
9700	323	313	350	338	369	356	381	367	384	370
9800	323	313	350	338	369	356	381	367	384	370
9900	323	313	350	339	369	356	381	367	384	370
10000	323	313	350	339	369	356	381	368	384	370
10100	323	313	350	339	369	356	381	368	384	370
10200	323	314	350	339	369	356	381	368	384	371
10300	323	314	350	339	369	357	381	368	384	371
10400	323	314	350	339	369	357	381	368	384	371
10500	323	314	350	339	369	357	381	368	384	371
10600	323	314	350	339	369	357	381	368	384	371
10700	323	314	350	339	369	357	381	368	384	371
10800	323	314	350	339	369	357	381	369	384	371
10900	323	314	350	340	369	357	381	369	384	371
11000	323	314	350	340	369	358	381	369	384	372
12000	323	315	350	340	369	358	381	370	384	373
13000	323	316	350	341	369	359	381	371	384	373
14000	323	316	350	342	369	360	381	371	384	374
15000	323	317	350	343	369	361	381	372	384	375
16000	323	317	350	343	369	361	381	373	384	375
17000	323	317	350	343	369	362	381	373	384	376
18000	323	318	350	344	369	362	381	374	384	376
19000	323	318	350	344	369	362	381	374	384	377
20000	323	318	350	344	369	363	381	374	384	377
21000	323	319	350	345	369	363	381	375	384	378
22000	323	319	350	345	369	363	381	375	384	378
23000	323	319	350	345	369	364	381	375	384	378
24000	323	319	350	345	369	364	381	376	384	378
25000	323	319	350	346	369	364	381	376	384	379
26000	323	319	350	346	369	364	381	376	384	379
27000	323	320	350	346	369	365	381	376	384	379
28000	323	320	350	346	369	365	381	376	384	379
29000	323	320	350	346	369	365	381	377	384	379
30000	323	320	350	346	369	365	381	377	384	380
40000	323	321	350	347	369	366	381	378	384	381
50000	323	321	350	348	369	367	381	379	384	381
60000	323	322	350	348	369	367	381	379	384	382
70000	323	322	350	349	369	368	381	379	384	382
80000	323	322	350	349	369	368	381	379	384	383
90000	323	322	350	349	369	368	381	380	384	383
100000	323	322	350	349	369	368	381	380	384	383
150000	323	323	350	350	369	369	381	381	384	384
200000	323	323	350	350	369	369	381	381	384	384
250000	323	323	350	350	369	369	381	381	384	384
300000	323	323	350	350	369	369	381	381	384	384

Fonte: EARLR / 1999.

ANEXO 05

DADOS DO TRE E IBGE

TABELA 22: Resultados Oficiais do TRE - Deputado Estadual - Ibirité -1998

Resultados Oficiais do TRE - Deputado Estadual

Sistema de Totalização - Módulo de Gerenciamento - [Consulta de Resultados]

Sistema Consulta Gerenciamento Configuração ?

FASE OFICIAL
Consulta de Resultados

Por UF/Brasil | Por Município | Por Zona

Município	**Cargo**	**Candidato**	
45950 IBIRITË	7-Deputado Estadual	99999	todos

Resumo | Gráfico

Candidato	Qtd.Votos	%Votos Válidos	%Comparecimento
41141 - Dinis Antônio Pinheiro	11544	31,02%	26,87%
23456 - Paulo Telles da Silva	6467	17,38%	15,05%
11554 - Antônio Genaro Oliveira	1064	2,86%	2,48%
41122 - João Paulo Gomes da Silva	762	2,05%	1,77%
12101 - Álvaro Antônio Teixeira Dias	498	1,34%	1,16%
13456 - João Bosco Senra	487	1,31%	1,13%
13139 - Durval Ângelo Andrade	411	1,10%	0,96%
22213 - Washington Fernando Rodrigues	398	1,07%	0,93%
25150 - Djalma Florêncio Diniz	329	0,88%	0,77%
13613 - Marco Antônio de Oliveira Zocrato	293	0,79%	0,68%
23133 - Wanderley Marques do Carmo	274	0,74%	0,64%
33123 - Adelino Carvalho Lino	268	0,72%	0,62%
45200 - João Leite da Silva Neto	249	0,67%	0,58%
23384 - Ronaldo Wagner Gontijo	243	0,65%	0,57%
33258 - Walter da Rocha Tosta	217	0,58%	0,51%
12212 - Alencar Magahães da Silveira Júnior	215	0,58%	0,50%
12140 - João Batista de Oliveira	206	0,55%	0,48%

Votos Brancos: 3902 (9,08%) Votos Nulos: 1848 (4,30%)
Seções Totalizadas: 156 (100,00%) Data/Hora da Totalização: 07/10/1998 10:36:20

MG - Minas Gerais | Consulta

Fonte: TRE

TABELA 23: Resultados Oficiais do TRE - Deputado Federal - Ibirité -1998

Resultados Oficiais do TRE - Deputado Federal

Sistema de Totalização - Módulo de Gerenciamento - [Consulta de Resultados]

Sistema Consulta Gerenciamento Configuração ?

FASE OFICIAL
Consulta de Resultados

Por UF/Brasil | Por Município | Por Zona

Município	**Cargo**	**Candidato**
45950 IBIRITË	7-Deputado Federal	99999 todos

Resumo | Gráfico

Candidato	Qtd.Votos	%Votos Válidos	%Comparecimento
2525 - Eliseu Resende	5000	13,57%	11,64%
1300 - Maria do Carmo Lara Perpétuo	3184	8,64%	7,41%
4544 - Vittorio Medioli	2829	7,68%	6,59%
1155 - Mário de Oliveira	1589	4,31%	3,70%
1514 - Luís do Nascimento	1410	3,83%	3,28%
2514 - José Perrela de Oliveira Costa	1386	3,76%	3,23%
1818 - Lincoln Diniz Portela	1357	3,68%	3,16%
2213 - Júlio César Gomes dos Santos	1288	3,50%	3,00%
4567 - Antônio Evangelista Teixeira	1232	3,34%	2,87%
4114 - Sérgio Antônio Mendes Figueiredo	1201	3,26%	2,80%
2345 - João Bosco Martins	733	1,99%	1,71%
1311 - Virgílio Guimarães de Paula	604	1,64%	1,41%
4555 - Aécio Neves da Cunha	574	1,56%	1,34%
1415 - Philemon Rodrigues da Silva	552	1,50%	1,28%
1111 - Herculano Anghinetti	466	1,26%	1,08%
4570 - Osmânio Pereira de Oliveira	367	1,00%	0,85%
4545 - Ademir Lucas Gomes	354	0,96%	0,82%

Votos Brancos: 4442 (10,34%) Votos Nulos: 1668 (3,88%)
Seções Totalizadas: 156 (100,00%) Data/Hora da Totalização: 07/10/1998 10:36:20

MG - Minas Gerais | Consulta

Fonte: TRE

TABELA 24: Resultados Oficiais do TRE - Presidente - Ibirité -1998

Resultados Oficiais do TRE Para Presidente

Tribunal Regional Eleitoral - TRE/MG | pág.: 1
Secretaria de Informática | 07/10/1998 16:53:56
Eleições Gerais de 1998 | 1º Turno
Gerenciamento - Versão 2.6 (Oficial) | GER210

Relatório Parcial do Município

Cargo: Presidente **Município**: 45950 - **IBIRITË**

Candidato	**Partido/ Coligação**	**Qtd. Votos**	**Ordem**	**%Vál**	**%Comp**
45 - Fernando Henrique Cardoso	PPB / PTB / PFL / PSD / PSDB	14941	1º	44.43%	34,78%
13 - Luiz Inácio Lula da Silva	PDT / PT / PCB / PSB / PC do B	14362	2º	42,71%	33,43%
23 - Ciro Ferreira Gomes	PL / PPS / PAN	3052	3º	9,08%	7,10%
56 - Enéias Ferreira Carneiro	PRONA	800	4º	2,38%	1,86%
33 - Ivan Moacyr da Frota	PMN	90	5º	0,27%	0,21%
19 - Thereza Tinajero Ruiz	PTN	86	6º	0,26%	0,20%
16 - José Maria de Almeida	PSTU	75	7º	0,22%	0,17%
43 - Alfredo Hélio Syrkis	PV	68	8º	0,20%	0,16%
31 - Vasco Azeredo Neto	PSN	59	9º	0,18%	0,14%
27 - José Maria Eymael	PSDC	39	10º	0,12%	0,09%
70 - João de Deus Barbosa de Jesus	PT do B	39	10º	0,12%	0,09%
20 - Sérgio Bueno	PSC	14	11º	0,04%	0,03%

Resumo	**Qtde.**	**%Comp**		**Qtde.**	**Percentual**
Votos Nominais:	33625	78,27%	Seções Totalizadas:	156	100,00%
Votos Brancos:	3282	7,64%	Comparecimento:	42961	84,64%
Votos Nulos:	6054	14,09%	Abstenção:	7799	15,36%

Fonte: TRE

TABELA 25: Resultados Oficiais do TRE - Governador - Ibirité -1998

Resultados Oficiais do TRE Para Governador

Tribunal Regional Eleitoral - TRE/MG — pág.: 1
Secretaria de Informática — 07/10/1998 16:54:52
Eleições Gerais de 1998 — 1º Turno
Gerenciamento - Versão 2.6 (Oficial) — GER210

Relatório Parcial do Município

Cargo: Governador **Município**: 45950 - **IBIRITÉ**

Candidato	**Partido/ Coligação**	**Qtd. Votos**	**Ordem**	**%Vál**	**%Comp**
15 - Itamar Augusto Cautiero Franco	PMDB / PSL / PST / PTN / PSC / PL / PPS / PAN / PRTB / PMN / PT do B	14716	1º	45,01%	34,25%
45 - Eduardo Brandão de Azeredo	PPB / PTB / PFL / PSN / PSD / PSDB	10114	2º	30,93%	23,54%
13 - Patrus Ananias de Sousa	PDT / PT / PCB / PSB / PV / PC do B	7649	3º	23,39%	17,80%
16 - Israel Pinheiro	PSTU	111	4º	0,34%	0,26%
44 - Milton Jun Yumiya	PRP	74	5º	0,23%	0,17%
36 - Francisco Simões de Aguiar	PRN	34	6º	0,10%	0,08%

Resumo	**Qtde.**	**%Comp**		**Qtde.**	**Percentual**
Votos Nominais:	33698	76,11%	Seções Totalizadas:	156	100,00%
Votos Brancos:	4427	10,30%	Comparecimento:	42961	84,64%
Votos Nulos:	5836	13,58%	Abstenção:	7799	15,36%

Fonte:TRE

TABELA 26: Distribuição Por Faixa Etária e Sexo da População de Ibirité

Município de Ibirité

	Total	*Homem*	*Mulher*
Faixa Etária	106781	53196	53585
Menos de 1 ano	2535	1210	1325
1ano	2528	1221	1307
2 anos	2566	1294	1272
3 anos	2575	1320	1255
4 anos	2403	1202	1201
5 anos	2430	1246	1184
6 anos	2509	1246	1263
7 anos	2521	1257	1264
8 anos	2481	1242	1239
9 anos	2343	1265	1078
10 anos	2360	1185	1175
11 anos	2209	1095	1114
12 anos	2437	1267	1170
13 anos	2469	1272	1197
14 anos	2559	1275	1284
15 a 19 anos	12183	6030	6153
20 a 24 anos	10635	5323	5312
25 a 29 anos	9866	5025	4841
30 a 34 anos	8940	4258	4682
35 a 39 anos	7906	3974	3932
40 a 44 anos	6068	3159	2909
45 a 49 anos	4166	2082	2084
50 a 54 anos	3046	1508	1538
55 a 59 anos	2328	1126	1202
60 a 64 anos	1654	782	872
65 a 69 anos	1194	509	685
70 a 74 anos	715	334	381
75 a 79 anos	406	163	243
80 anos e mais	353	125	228
Idade Ignorada	396	201	195

Fonte : Fundação Instituto Brasileiro de Geografia e Estatística - IBGE

TABELA 27: Pessoas Residentes de 15 anos ou mais de Idade por Anos de Estudo Segundo o Sexo e os Grupos de Idade

Sexo e Grupo de Idade	Sem Instrução e menos de 1 ano	1 a 3 anos	4 a 7 anos	8 a 10 anos	11 a 14 anos	15 anos ou mais
Homem						
15 a 19	207	1086	4513	1112	105	0
20 a 24	459	865	3861	1238	497	8
25 a 29	208	872	3223	1008	471	19
30 a 34	289	950	2286	770	320	21
35 a 39	335	1075	2464	482	226	11
40 a 44	336	995	1797	400	136	12
45 a 49	338	737	1119	188	89	17
50 a 54	335	838	663	92	49	8
55 a 59	352	484	419	52	31	4
60 a 64	318	349	236	29	17	4
65 anos ou mais	655	451	288	48	18	5
Mulher						
15 a 19	155	804	4389	1700	218	0
20 a 24	156	755	3165	1268	830	6
25 a 29	214	546	3042	928	815	35
30 a 34	291	1076	2919	721	425	32
35 a 39	365	1154	2332	439	230	24
40 a 44	442	976	1569	295	139	20
45 a 49	583	765	906	130	84	15
50 a 54	511	630	545	71	39	11
55 a 59	551	453	346	37	23	5
60 a 64	529	354	183	20	16	2
65 anos ou mais	1177	445	238	26	26	1

Fonte: Fundação Instituto Brasileiro de Geografia e Estatística - IBGE

ANEXO 06

TRAÇADO URBANO DE IBIRITÉ

ANEXO 07

ADAPTAÇÃO DE PROJETOS DE AMOSTRAGEM

7.1 AMOSTRAGENS ADAPTADORAS

7.1.1 Coleta, Recoleta de Amostra e Estimadores

Na coleta e recoleta de amostra para se estimar o número total de indivíduos de uma população, obtém-se uma amostra aleatória inicial, e os indivíduos dessa amostra são marcados ou identificados. Uma segunda amostra é independentemente obtida e o número de indivíduos marcados é observado. Se a segunda amostra for representativa da população como um todo, ela deve conter uma proporção próxima da proporção de indivíduos marcados da população. Dessa relação, pode-se estimar o total de indivíduos da população. Os métodos de coleta e recoleta têm sido usados para se estimar a população de animais (aves, mamíferos, peixes, mosquitos e outros), como também para estimar o contingente populacional elusivo e para ajustar as subestimativas do censo dos grupos minoritários, estimando-se o número de eventos vitais em uma população.

Notação:

$\Re$ total de indivíduos da população.

X número de indivíduos marcados na primeira amostra

y total de indivíduos da segunda amostra

x número de indivíduos marcados da segunda amostra

Uma amostra aleatória de tamanho "n" refere-se a uma seleção de "n" unidades retiradas de gráficos, seções, trilhas, regiões de atração, linhas através de uma seção, mapas de variável circular ou outras unidades de detectabilidade, pelas quais os indivíduos de uma população são observados.

O total de elementos $\Re$ da população pode então ser estimado, supondo-se que a proporção de elementos marcados da 2ª amostra é representativa da proporção de elementos marcados da população. Assim,

$$\frac{x}{y} = \frac{X}{\Re} \text{ , então} \tag{7.1.1.1}$$

e a solução para o tamanho da população total (desconhecido) $\Re$ tem o estimador de Petersen

$$\hat{r} = \frac{y}{x} X \ . \tag{7.1.1.2}$$

Um estimador da variância de $\hat{r}$ (Sekar e Deming 1949) é

$$S^2(\hat{r}) = \frac{Xy(X-x)(y-x)}{x^3} \tag{7.4.1.3}$$

Um intervalo de confiança de 100(1-α)% para $\Re$ é dado por:

$$\hat{r} \pm Z\sqrt{S^2(\hat{r})} \tag{7.1.1.4}$$

O número x de elementos marcados para a 2ª amostra pode ser zero. A variância não será finita e os estimadores não tendenciosos (propostos por Chapman) para o total da população e para a variância são:

$$\bar{r} = \frac{(x+1)(y+1)}{(x+1)} - 1 \quad \text{ e} \tag{7.1.1.5}$$

$$S^2(\bar{r}) = \frac{(x+1)(y+1)(X-x)(y-x)}{(x+1)^2(x+2)} \ . \tag{7.1.1.6}$$

Para um modelo Multinomial Geral, a coleta ou detecção de qualquer elemento da população pode ser classificada em quatro categorias (detecção na 1ª ou 2ª ocasiões, detecção na 1ª ocasião, detecção na 2ª ocasião e não detecção). Assim, um modelo multinomial com as quatro probabilidades para as quatro células que se acrescentam a uma única célula, aplica-se ao procedimento de coleta de cada elemento. Para este modelo, o estimador de máximoverossimilhança para a população total é representado pela parte inteira do estimador de Petersen $\hat{r} = X\dfrac{y}{x}$, na qual a probabilidade de coleta da 1ª amostra $\hat{p}_1 = x/y$ e da 2ª amostra é $\hat{p}_2 = x/X$. Mesmo que as probabilidades de coleta sejam diferentes para diferentes indivíduos na 1ª amostra, mas igual à 2ª, o estimador será ainda o de Petersen. Se uma probabilidade p se aplicar tanto entre as amostras como a todos os indivíduos, com independência entre as amostras, os estimadores de máximo-verossimilhança para p e $\Re$ serão respectivamente,

$$\hat{p} = 2x/(X+y) \quad \text{e} \quad \hat{r} = (X+y)/2\hat{p} \tag{7.1.1.7}$$

Se os valores de X e y para as duas amostras forem fixos, e a 2ª amostra for uma a.a. sem reposição, então o número x de elementos marcados terá distribuição hipergeométrica (Thompson, 1992).

Sob o modelo de Poisson, se o número das categorias para a coleta forem supostamente variáveis aleatórias independentes, o número total $\Re$ de indivíduos da população também será aleatório.

Para uma a.a. simples (com reposição) de n trajetos, mapas, ou onde outros tipos de unidades são selecionados, seja a i'ésima unidade da amostra, em que y_i representa o número de elementos observados, e x_i o número de elementos marcados, então, o número total de elementos para a amostra é $\sum_{i=1}^{n} y_i$, dos quais $\sum_{i=1}^{n} x_i$ são marcados. As estatísticas $\bar{x}$ e $\bar{y}$

representam respectivamente as médias da amostra. O total X de elementos marcados da população é conhecido, quando se supõe uma população fechada.

Seja o número marcado e liberado na 1ª ocasião de amostragem, a estimativa de Petersen para o número de elementos da população é:

$$\hat{r} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} X = \frac{\sum_{i=1}^{n} \frac{y_i}{n}}{\sum_{i=1}^{n} \frac{x_i}{n}} = \frac{\bar{y}}{\bar{x}} X \; . \qquad (7.1.1.8)$$

Observa-se que a estimativa acima nada mais é que a estimativa de um índice clássico, e a estimativa da variância será:

$$S^2(\hat{r}) = (\frac{X}{\bar{x}})^2 \frac{1}{(n(n-1))} \sum_{i=1}^{n} (y_i - rx_i)^2 \quad \text{onde} \quad r = \frac{\bar{y}}{\bar{x}} \; . \qquad (7.1.1.9)$$

Um intervalo de confiança de 100(1-α)% para $\Re$ será dado por:

$$\hat{r} \pm t\sqrt{S^2(\hat{r})} \; , \qquad (7.1.1.10)$$

onde $t = t_{n-1;a/2}$ da distribuição de Student[19] e quando $n \rightarrow \infty$ pode-se substituir o valor de t pelo valor de Z (N(0;1)).

Se os elementos marcados estiverem mal distribuídos na região, de tal forma que cada unidade x_i possa ser igual a zero, para algumas amostras, então o estimador proporcional, como o usual de Petersen, não terá uma variância finita.

A amostragem aleatória sem reposição será possível apenas se os elementos da população forem agrupados em unidades não sobrepostas para seleção, ou quando uma região de estudo espacial for dividida em unidade de área com probabilidade constante de detecção para todos os indivíduos da unidade selecionada. Coleta e recoleta em que a 2ª amostra

---

[19] Veja anexo 03.

consista em uma a.a. de n unidades sem reposição, de uma população de N elementos, e a probabilidade de captura ou detecção seja constante para todos os indivíduos de uma unidade amostral são descritas por Seber (1982) e Wolter (1986). O estimador para a população total $\Re$ é o estimador de Petersen, e o estimador da variância fornecido por Wolter é

$$S^2(\hat{r}) = (\frac{X}{\bar{x}})^2 \frac{(N-n)}{Nn(n-1)} \sum_{i=1}^{n} (y_i - rx_i)^2 + \frac{X\bar{y}(m_x - \bar{x})(\bar{y} - \bar{x})}{\bar{x}^3}, \tag{7.1.1.11}$$

onde $m_x = X / N$. Wolter coloca o teorema do limite central para distribuição normalmente assintótica do estimador de Petersen sob amostragem aleatória simples sem reposição. Estimar o tamanho da população, explícita ou implicitamente, resulta em estimativas ou probabilidades de detecção de coleta. O estimador de Petersen tem a forma em comum com a de muitos outros estimadores para grandes populações ou densidade em que a detectabilidade seja imperfeita,

$$\hat{r} = \frac{y}{\hat{p}}, \tag{7.1.1.12}$$

onde $\hat{p} = x / X$ que estima a probabilidade de detecção da 2ª amostra. Se a distribuição de x for a distribuição hipergeométrica, a variância da detectabilidade estimada, dados X e y, será

$$Var(\hat{p}) = \frac{y}{X^2}(\frac{X}{r})(1-\frac{X}{r})(\frac{r-y}{r-1}). \tag{7.1.1.13}$$

Uma estimativa da detectabilidade por unidade é

$$\hat{p} = \bar{x} / X \quad \text{onde } \bar{x} = \sum_{i=1}^{n} x_i . \tag{7.1.1.14}$$

A estimativa da variância é

$$S^2(\hat{p}) = s_x^2 /(X^2 n) \text{ onde } s_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 /(n-1) . \tag{7.1.1.15}$$

7.1.2 Amostragem Interceptora de Linha

Na amostragem de linha, uma amostra aleatória de linhas é selecionada numa área de estudo e, sempre que um objeto da população é interceptado por uma ou mais linhas da amostra, uma variável de interesse associada ao objeto é registrada. Um exemplo desse tipo de amostragem é o estudo do habitat ecológico, em que o objeto de estudo é a estimativa da quantidade total de frutos de uma certa espécie de plantas em uma área de estudo. Uma a.a. de n linhas para cada um dos tamanhos L é selecionada e registrada no mapa da área de estudo. Os pesquisadores de campo percorrem cada linha e toda vez que a linha interceptar um conjunto das espécies de interesse, os frutos desse conjunto serão coletados e sua quantidade (Y_k) será avaliada. Com esse esquema de amostragem, um conjunto grande de árvores tem maior probabilidade de inclusão na amostra que um conjunto pequeno, e uma estimativa não tendenciosa das quantidades da população depende da determinação destas probabilidades (Thompson, 1992).

O esquema de amostragem de linhas com direção fixada é o mais simples. Neste esquema de amostragem, "n" linhas transversais são selecionadas aleatoriamente, selecionando-se "n" posições ao longo de uma linha básica de extensão b, que corta perpendicularmente a largura da região de estudo e percorre uma transversal perpendicular através da área de estudo para a linha de base, em cada uma das posições selecionadas. Se K denota o número de objetos da população associado ao k'ésimo objeto, então esta é uma variável de interesse Y_k. O objetivo é estimar a população total onde $r = \sum_{i=1}^{K} Y_i$ ou a densidade por área de unidade D=r/A, onde A é a área da região de estudo.

Para qualquer estrutura dada, a probabilidade de que a linha selecionada intercepte o k'ésimo objeto é proporcional à largura w_k ao longo da linha de base do conjunto de pontos, para a qual a perpendicular intercepta o objeto de ordem k. Assim, w_k é a largura de alcance da sombra do objeto k sobre a linha de base. A probabilidade de seleção estrutura por estrutura é dada por:

$$P_k = \frac{w_k}{b} . \qquad (7.1.2.1)$$

Se C_i for o conjunto de objetos da população interceptados pela i'ésima linha transversal da amostra, para cada um dos objetos interceptados, divide-se o valor da variável de interesse Y_k pela probabilidade de seleção P_k e se define uma nova variável v_i como sendo a soma

$$v_i = \sum_{k \in C_i} \frac{Y_k}{P_k} . \qquad (7.1.2.2)$$

A variável v_i é um estimador não tendencioso da população total $\Re$, e a amostra aleatória de n linhas transversais fornece v_1, v_2, ..., v_n, que são independentes e identicamente distribuídos (i.i.d).

A média da amostra é

$$\hat{r}_p = \frac{1}{n} \sum_{i=1}^{n} v_i , \qquad (7.1.2.3)$$

que representa um estimador não tendencioso de $\Re$, com variância $V(\hat{r}_p) = \frac{1}{n} V(v_i)$. O estimador de $\hat{r}_p$ é semelhante ao estimador de Hansen-Hurwitz. Na presente estrutura, entretanto, os objetos não são selecionados independentemente, uma vez que ocorrem seleções conjuntas com linhas que interceptam o mesmo objeto. Supondo-se que $S^2(v)$ represente a variância da amostra dos v'is, então,

$$S_v^2 = \frac{1}{n-1}\sum_{i=1}^{n}(v_i - \hat{r}_p)^2 \ . \qquad (7.1.2.4)$$

Assim, um estimador não tendencioso da variância de $\hat{r}_p$ é dado por:

$$S^2(\hat{r}_p) = \frac{S^2{}_v}{n} \qquad (7.1.2.5)$$

Não há fator de correção de população finita (cpf) porque a seleção de posições ao longo da linha de base é essencialmente com reposição. Mesmo com posições transversais distintas, um objeto dado pode ser interceptado por mais de uma linha transversal, e daí ser contado mais de uma vez no estimador $\hat{r}_p$. Uma estimativa que dependa apenas da interseção de objetos distintos, interceptados pela amostra de linhas transversais, pode ser obtida pelo método de Horvitz-Thompson. Suponhamos que k seja o número de objetos distintos interceptados.

A probabilidade de que o k'ésimo objeto esteja incluído na amostra é

$$p_k = 1-(1-P_k)^n \ . \qquad (7.1.2.6)$$

O estimador de Horvitz-Thompson é

$$\hat{r}_p = \sum_{k=1}^{K}\frac{Y_k}{p_k} \ . \qquad (7.1.2.7)$$

7.2 Adaptação de Projetos de Amostragem

Na teoria de amostragem e na metodologia, as atenções foram direcionadas para o projeto de amostragem, no qual o procedimento de seleção não depende de qualquer modo das observações realizadas durante o levantamento, de modo que as unidades amostrais possam ser determinadas antes do levantamento. Em muitas situações, o pesquisador pode se sentir inclinado a tomar decisões durante um levantamento, baseado no que foi observado até o momento. Essas decisões referem-se a quais áreas e quantas áreas devam ser observadas para o término do experimento (Thompson, 1992).

A adaptação de amostragem é colocada para os projetos de amostragem nos quais o procedimento para seleção de áreas ou unidades, a serem incluídas na amostra, dependam das observações encontradas das variáveis de interesse durante o levantamento.

A finalidade principal dos projetos de adaptação de amostragem é tirar vantagem das características da população, de tal forma que as estimativas da abundância ou densidade da população, para uma quantidade ou custo, sejam mais precisas. Geralmente, a localização e a forma dos agrupamentos não podem ser previsíveis antes de um levantamento. Os meios tradicionais de aumento de precisão, tais como a estratificação, não são suficientes. Para tais populações, estratégias de adaptação de amostragem podem aumentar sistematicamente a precisão e resultar em melhores estimativas de outros parâmetros de interesse.

Procedimentos de adaptação de seleção podem introduzir tendências nos estimadores convencionais, de modo que os estimadores que são não tendenciosos sob os projetos de adaptação são necessários e a não tendenciosidade não dependerá das suposições populacionais. Os métodos de adaptação de amostragem fornecem grandes aumentos em eficiência para algumas populações, mesmo que os meios convencionais de aumento de precisão (estratificação, arranjo sistemático ou de conglomerado) tenham sido aplicados. Para uma dada população, a escolha definitiva do procedimento de seleção depende tanto das características conhecidas da população quanto das características práticas, como o custo e conveniências" (Thompson, 1992)[20].

7.2.1 Adaptação de Amostragem Por Conglomerado

"Os estimadores desenvolvidos para amostragem adaptadora por conglomerado estão relacionados com os estimadores de Hansen-Hurwitz e Horvitz-Thompson. Na amostragem adaptadora de conglomerado, as probabilidades de seleção e inclusão necessárias para os estimadores não podem ser determinadas para cada unidade da amostra e, assim, os estimadores são modificados.

Como em amostragem de população finita, a população consiste em N unidades numeradas 1, 2, ... , N associadas com variáveis de interesse $y=(y_1, y_2, ..., y_N)$. A amostra S é um conjunto ou seqüência de números identificadores das unidades selecionadas para observação. Os dados consistem de y valores observados associados, com os números selecionados. O objeto de interesse é estimar a média da população

$$m = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad (7.2.1.1)$$

ou o total $t = Nm$ dos y valores. Um projeto de amostragem é uma função P(S/y) que determina uma probabilidade para toda amostra S possível, e essas probabilidades de seleção dependem dos y valores da população.

[20] Tradução livre.

Supõe-se que para toda unidade i da população, uma vizinhança A_i seja definida de uma coleção de unidades que inclua i. Essas vizinhanças não dependem dos y valores da população. Em amostragem espacial, a vizinhança de cada unidade consiste de um conjunto de vizinhos geograficamente mais próximos com padrões de vizinhança mais elaborados, em que se inclui um conjunto (contíguo ou não) maior de unidades, como um padrão de rede (grade) sistemático ao redor da unidade inicial. Em outros casos, as vizinhanças podem ser definidas pelos relacionamentos sociais ou institucionais entre as unidades. A relação de vizinhança será simétrica se a unidade j estiver na vizinhança da unidade i, então, a unidade i está na vizinhança da unidade j.

A condição das unidades da vizinhança é dada por um intervalo ou conjunto C na série da variável de interesse. A unidade i é considerada satisfatória da condição, se $y_i \in C$. Uma unidade satisfaz a condição se a variável de interesse y_i for maior ou igual a alguma constante c, isto é, c = (y: y≥c). Quando uma unidade selecionada satisfaz a condição, todas as unidades dentro de sua vizinhança são acrescentadas à amostra e observadas. Algumas dessas unidades podem, por sua vez, satisfazer ou não a condição. Para qualquer dessas unidades que satisfaça a condição, as unidades em sua vizinhança são também incluídas na amostra, e assim por diante.

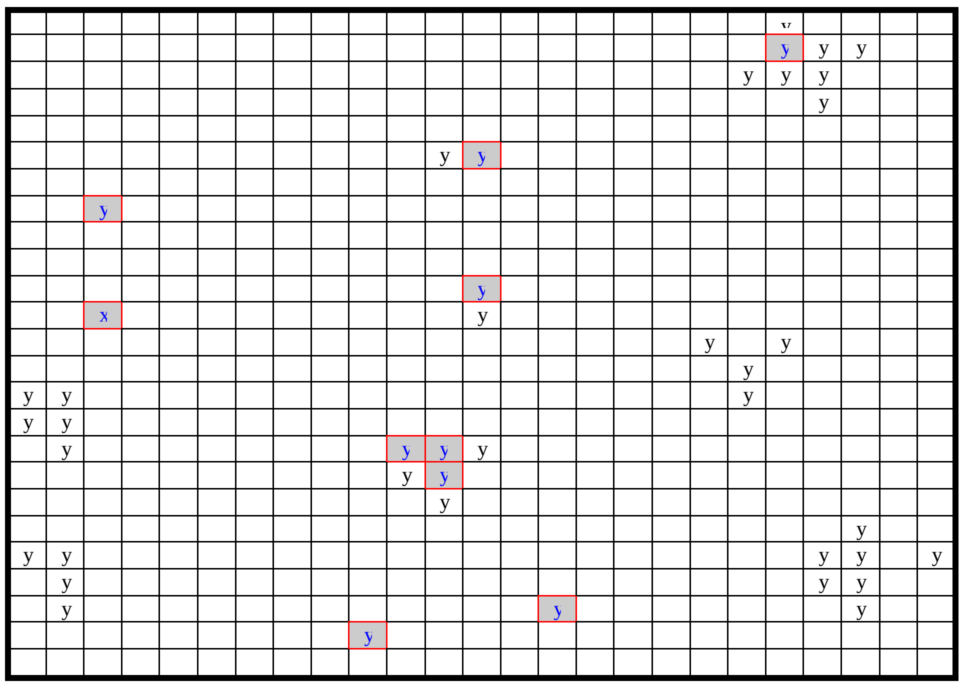
Considere o conjunto de todas as unidades que são observadas sob o projeto como um resultado da seleção inicial da unidade i. Tal seleção pode consistir da união de várias vizinhanças e será denominada um conglomerado quando aparecer numa amostra. Dentro de tal conglomerado está, um subconjunto de unidades denominado rede com a propriedade de que a seleção de qualquer unidade dentro da rede conduzirá à inclusão na amostra de unidades alternadas. Dentro de cada um dos conglomerados de unidades da amostra final, o subconjunto de unidades com um ou mais pontos dos objetos formam uma rede. Qualquer objeto que não satisfaça a condição, exceto na vizinhança de uma que satisfaça, é denominado unidade de limite. Enquanto a seleção de qualquer unidade da rede resultar em inclusão de todas as unidades de limite associadas, a seleção de uma unidade de limite não resultará na inclusão de quaisquer outras unidades. É conveniente considerar qualquer unidade que não satisfaça a condição numa rede de dimensão unitária, de modo que, dados os y valores, a população possa ser dividida exclusivamente em redes (Thompson, 1992).

Para exemplificar este tipo de amostragem, pode-se observar as figuras 30 e 31 . A figura 30 representa os pontos (y) para uma a.a.simples sem a inclusão de elementos na amostra, e a figura 31 representa a amostra final, em que se incluem os elementos das vizinhanças.

Se a amostra inicial de n unidades for obtida através da amostragem simples sem reposição, as n unidades da amostra inicial serão distintas devido ao tipo de amostragem, mas os dados poderão conter observações repetidas, devido à seleção da unidade inicial de mais de uma unidade no conglomerado inicial. A unidade i será incluída na amostra, se qualquer unidade da rede à qual ela pertence (incluindo ela mesma) for selecionada como parte da amostra inicial, ou se qualquer unidade de uma rede da qual a unidade i, que é uma unidade de limite, for selecionada. Suponha que a_i represente o número total de unidades em redes cuja unidade i seja uma unidade de limite. Se a unidade i satisfizer o critério C, então $a_i = 0$, enquanto que, se a unidade i não satisfizer a condição, então $m_i = 1$. A probabilidade de que a unidade i seja selecionada em qualquer uma das extrações é $pi = (m_i + a_i)/N$, e a probabilidade de que a unidade i seja incluída na amostra é:

$$p_i = \binom{\mathbf{N} - \mathbf{m_i} - a_i}{n} \qquad (7.2.1.2)$$

FIGURA 30: Amostragem Simples

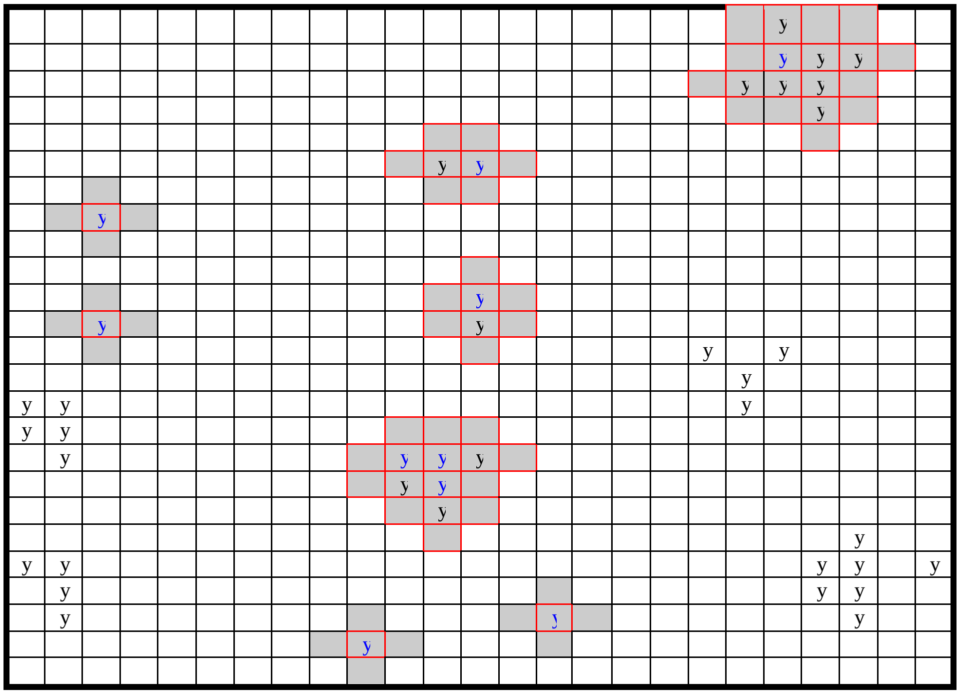


Se a amostra inicial for simples, mas com reposição, as observações repetidas dos dados podem ocorrer devido às seleções repetidas da amostra inicial ou à seleção inicial de mais de uma unidade de um conglomerado. Com esse projeto de amostragem, a probabilidade de seleção, ponto a ponto, é $p_i = (m_i + a_i)/N$ e a probabilidade de inclusão é

$$p_i = 1 - (1 - p_i)^n . \qquad (7.2.1.3)$$

Para cada um dos projetos iniciais de seleção (com ou sem reposição), as probabilidades de seleção e de inclusão não podem ser determinadas a partir dos dados para todas as unidades da amostra, pois alguns dos a_i podem ser desconhecidos.

FIGURA 31: Amostragem Simples com a Inclusão de Elementos



Os estimadores não tendenciosos da média populacional para amostragem aleatória simples $(\bar{y})$ e para amostragem por conglomerados $(\bar{\bar{y}})$, já conhecidos, tornam-se tendenciosos quando se usa o projeto adaptador de amostragem. Se a amostra inicial for selecionada pela a.a. simples (com ou sem reposição), a média da amostra $(\bar{y})$ das n observações iniciais como estimador levará em conta apenas as observações iniciais, tornando-se, portanto, tendenciosa.

Para projetos de amostras em que n unidades são selecionadas com reposição, e a probabilidade p_i de se selecionar a unidade i em qualquer retirada seja conhecida para todas as unidades, o estimador de Hansen-Hurwitz, em que cada valor de y é dividido pela probabilidade de seleção associada e multiplicado pelo número de vezes que a unidade for selecionada, é um estimador não tendencioso da média populacional. Como nos projetos adaptadores para conglomerados, as probabilidades de seleção não são conhecidas, um estimador não tendencioso pode ser fornecido pela modificação do estimador de Hansen-Hurwitz, de tal forma que as observações não iniciais façam parte da estimativa. Assim, o estimador modificado será baseado nas probabilidades de retirada (ponto a ponto) de uma rede de unidade cujo interesse seja a amostra inicial.

Seja y_i a rede que inclui a unidade i e m_i o número de unidades nessa rede, um estimador modificado baseado nas probabilidades de retirada de uma unidade de rede faça parte da amostra inicial é:

$$w_i = \frac{1}{m_i} \sum_{j \in y_i} y_j \tag{7.2.1.4}$$

e o estimador modificado será dado por

$$\bar{y}_1 = \frac{1}{n} \sum_{i=1}^{n} w_i \tag{7.2.1.5}$$

A variância de $\bar{y}_1$, se a amostra for selecionada sem reposição, será dada por

$$Var(\bar{y}_1) = \frac{(N-n)}{Nn(N-1)} \sum_{i=1}^{N} (w_i - m)^2 \tag{7.2.1.6}$$

e, se a amostra inicial for selecionada com reposição:

$$Var(\bar{y}_1) = \frac{1}{n} \sum_{i=1}^{N} (w_i - m)^2 / N \tag{7.2.1.7}$$

Os estimadores não tendenciosos para a $Var(\bar{y}_1)$ para amostras retiradas sem e com reposição respectivamente, serão dados por:

$$S^2(\bar{y}_1) = \frac{(N-n)}{Nn(n-1)} \sum_{i=1}^{n} (w_i - \bar{y}_1)^2 \quad \text{e} \tag{7.2.1.8}$$

$$S^2(\bar{y}_1) = \frac{1}{(n(n-1)}\sum_{i=1}^{n}(w_i - \bar{y}_1)^2 \tag{7.2.1.9}$$

A não tendenciosidade dos projetos adaptadores não depende do tipo de população amostrada, pois é baseada no próprio projeto. Mas a eficiência de um projeto de amostragem depende do tipo de população amostrada. Ao se comparar a variância de uma população finita para a média de uma amostra aleatória simples adaptadora de conglomerado de n unidades sem reposição, com a variância de uma amostra aleatória simples não adaptadora, o projeto adaptador fornecerá variância menor do que a variância da média da a.a. simples de tamanho n* se e somente se:

$$(\frac{1}{n} - \frac{1}{n^*})s^2 < \frac{N-n}{Nn(N-1)}\sum_{k=1}^{K}\sum_{i\in y_k}(y_i - w_i)^2 \tag{7.2.1.10}$$

onde:

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - m)^2 \tag{7.2.1.11}$$

e y_k é a k'ésima rede da população. Observa-se que, na desigualdade, o termo à direita contém a variância interna da rede da população. Desse modo, a amostragem adaptadora de conglomerado com o estimador $\bar{y}_1$ será mais eficiente do que a amostragem aleatória simples, se a variância interna da rede da população for suficientemente elevada"(Thompson, 1992)[21].

7.2.1 Amostragem Adaptadora de Conglomerado e Sistemática por Faixa

"Os projetos de amostragem adaptadora de conglomerado são projetos nos quais a amostra inicial é selecionada em termos de unidades primárias, e adições subseqüentes à amostra são em termos de unidades secundárias. Por exemplo, num levantamento aéreo em que as unidades pesquisadas sejam os leões marinhos ou os ursos polares, ou num levantamento marítimo, em que a unidade de estudo sejam as baleias, se a faixa observada em cada transversal selecionada formar uma unidade primária e se toda vez em que os animais forem avistados, a área ao lado da transversal for pesquisada, ou ainda com levantamento posterior, se os animais adicionais forem identificados enquanto estiverem na pesquisa, o padrão da pesquisa continuará a definir as vizinhanças das unidades secundárias acrescentadas à amostra.

"Nas observações de aves e espécies de peixes, a seleção das áreas em que as observações serão feitas é sempre realizada sistematicamente, e uma seleção simples sistemática forma uma unidade primária. Se observações adicionais forem feitas na vizinhança de qualquer área em que haja abundância, as observações subseqüentes não seguirão, geralmente, o padrão sistemático inicial. Com tais situações de levantamento, pode-se pensar na região de estudo como se estivesse subdividida em unidades secundárias, representando todas as áreas possíveis em que as observações possam ser realizadas, enquanto as unidades primárias, das quais a amostra inicial foi selecionada, consistem de conglomerados, tais como faixas longas e estreitas ou arranjos sistemáticos das unidades secundárias.

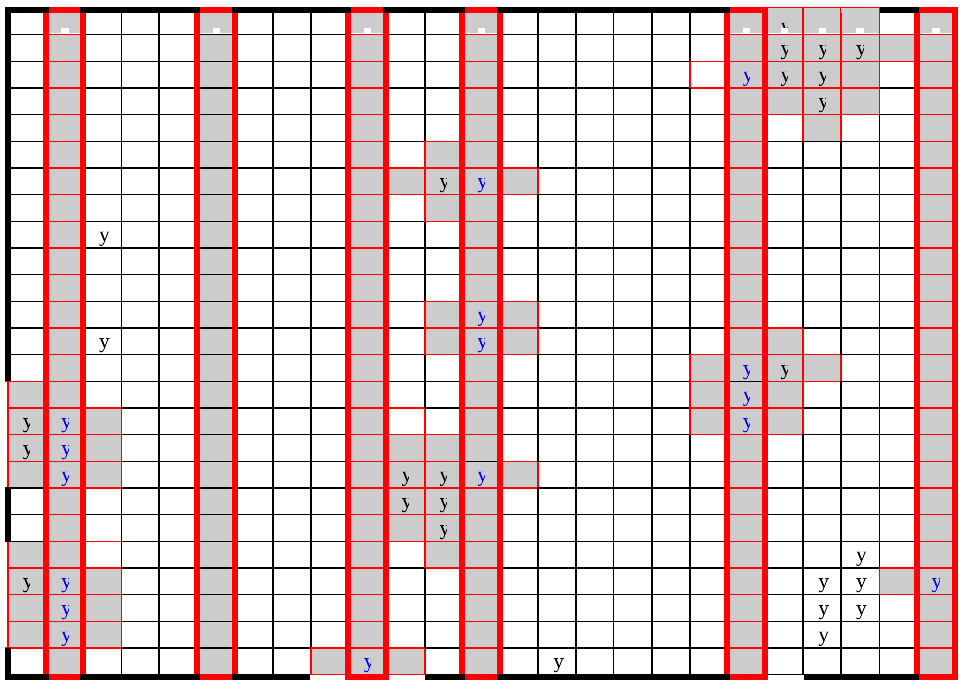
As figuras 32 e 33 exemplificam estes tipos de projeto de amostragem, nos quais o objetivo é estimar o número médio de elementos de uma população na região de estudo. Observa-se a figura 32, em que a amostra inicial consiste de seis faixas selecionadas aleatoriamente (unidades primárias) e as unidades secundárias são áreas pequenas e quadradas. Toda vez que uma área de amostra contiver um ou mais elementos com as características em estudo, as áreas adjacentes serão adicionadas à amostra. Se por sua vez qualquer dos novos

---

[21] Tradução livre.

locais da amostra contiver pelo menos um elemento com as características, as áreas adjacentes adicionais (sentidos norte-sul e leste-oeste) serão adicionadas na amostra.

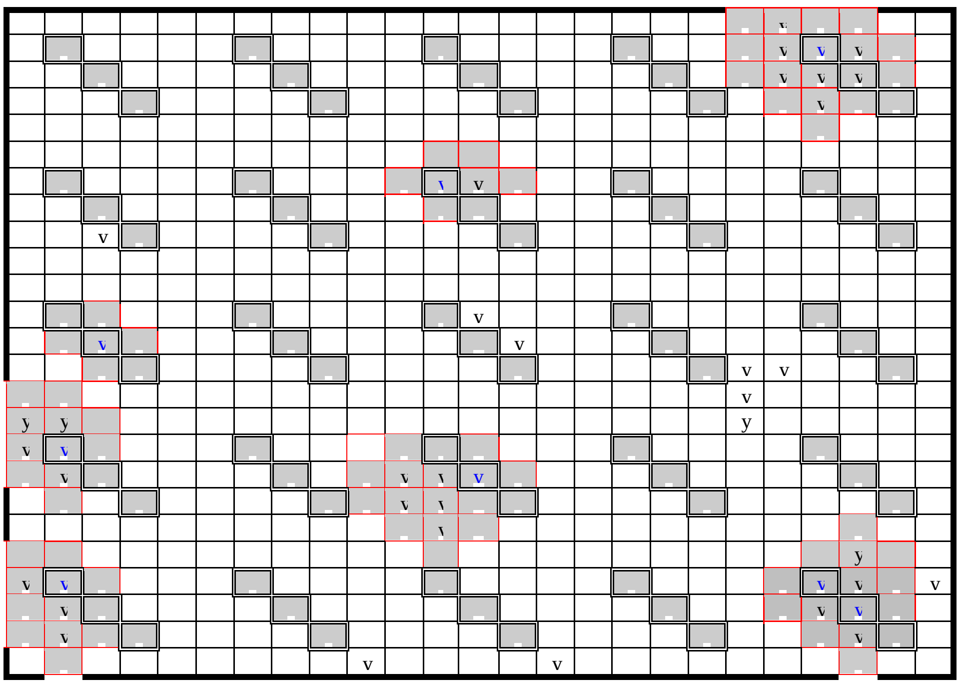
FIGURA 32: Amostragem por Conglomerado



Na figura 33, a amostra inicial é uma amostra sistemática espacial com três pontos de partida selecionados aleatoriamente. Toda vez que observações com as características em estudo são encontradas em qualquer área da amostra, as áreas adjacentes são acrescentadas à mesma.

Para os projetos de amostragem adaptadora de conglomerado, a população é composta de N unidades primárias, em que cada uma contém m unidades secundárias. As N.M unidades da população são denotadas por u_{ij} , para i=1, 2, ..., N e j=1, 2, ..., M. Associada à j'ésima unidade secundária da i'ésima unidade primária está a variável de interesse y_{ij}. O interesse está na estimação da média populacional $m = \frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}$ ou equivalentemente da população total $\Re = MNm$. (7.2.2.1)

FIGURA 33: Amostragem Sistemática



Para toda unidade secundária da população, uma coleção de unidades, chamada vizinhança dessa unidade, é definida. A vizinhança da unidade u_{ij} inclui a unidade u_{ij} e, se a unidade u_{ij} pertencer à vizinhança da unidade $u_{i'j'}$, então a unidade $u_{i'j'}$ pertencerá à vizinhança da unidade uij. Em aplicações, a vizinhança da unidade é definida como um conjunto contíguo das unidades circunvizinhas ou como um padrão sistemático das unidades circunvizinhas.

Nos exemplos das figuras 32 e 33, as unidades secundárias são representadas pelas áreas quadradas, e as vizinhanças dessas áreas consistem em si mesmas mais suas unidades adjacentes nos sentidos norte-sul e leste-oeste, de modo que para uma área fora dos limites da área de estudo, a vizinhança consista de cinco áreas em forma de cruz. Muitas outras configurações vicinais são possíveis. Uma vizinhança pode, de fato, consistir de um conjunto de unidades não contíguas, espalhadas como num padrão de grade sistemático sobre a unidade original.

A unidade u_{ij} satisfaz a condição de interesse, se o valor associado y_{ij} estiver num conjunto especificado C. Para problemas relacionados à quantidade de elementos com a característica em estudo, a condição pode comumente ser definida de modo que a unidade satisfaça a condição de que seu valor y seja igual ou maior que alguma constante c.

Nos projetos adaptadores de amostragem de conglomerado, uma a.a. de n unidades primárias é selecionada por amostragem aleatória simples sem reposição. Sempre que um valor observado de unidade secundária da amostra satisfizer a condição de interesse, todas as unidades em sua vizinhança serão incluídas na amostra. Por sua vez, se qualquer dessas unidades adicionais subseqüentemente satisfizerem a condição, as unidades de sua vizinhança serão também adicionadas à amostra, de modo que a amostra contenha toda unidade da vizinhança de qualquer unidade da amostra que satisfaça a condição.

Uma população com um conjunto dado de y valores pode ser subdividida em K conjuntos denominados por rede, de modo que sempre que uma unidade u_{ij} que satisfizer a condição, estiver na vizinhança da unidade $u_{i'j'}$, que também satisfaça a condição, então as unidades u_{ij} e $u_{i'j'}$ pertencerão à mesma rede. Assim, se uma unidade primária, inicialmente selecionada, interceptar uma rede dada, toda unidade dessa rede estará incluída nessa amostra. Uma unidade que não satisfizer a condição representará uma rede que consiste apenas de si mesma. Nas figuras 32 e 33, observam-se redes que são maiores que o tamanho da área simples (inicial), e cada área dentro de uma dessas grandes redes contém, pelo menos, um elemento com as características em estudo. O símbolo utilizado para representar uma ou mais unidades com as características em estudo nas figuras 32 e 33 foi a letra y.

Uma unidade u_{ij} que não satisfizer a condição, será incluída na amostra, ou se a unidade primária que a inclui for selecionada inicialmente, ou se qualquer unidade primária selecionada inicialmente interceptar a rede e satisfizer a condição de vizinhança da unidade u_{ij}. Uma unidade que não satisfizer a condição, mas estiver na vizinhança de uma ou mais unidades que satisfaçam a condição, será denominada unidade de limite, enquanto as vizinhanças serão definidas por relações como a proximidade física e não dependerão dos valores de y da população. As redes dependem dos valores y da população que correspondem às agregações naturais dos indivíduos da população (animais, plantas e outros). Se cada unidade inicial consistir de um conjunto de unidades da população, dispostas uniformemente, a amostra inicial deverá ser a sistemática, conforme a figura 33. As unidades primárias iniciais serão chamadas de faixas se cada unidade primária inicial consistir de uma fila de unidades dispostas em uma linha reta. Existem outras formas de unidades primárias (Thompson, 1992).

A probabilidade de seleção ponto a ponto é dada por:

$$p_{ij} = \frac{m_{ij} + a_{ij}}{N} \qquad (7.2.2.2)$$

onde, m_{ij} é o número de unidades primárias que interceptam a rede que contém a unidade u_{ij}, e a_{ij} é o número de unidades primárias que não interceptam a rede da unidade u_{ij}, mas interceptam a rede de uma ou mais unidades que satisfazem a condição na vizinhança da unidade u_{ij}. Para uma unidade que satisfaça a condição, $a_{ij} = 0$, enquanto que, para uma unidade que não satisfaça a condição, $m_{ij} = 1$.

A probabilidade p_{ij} de que a unidade u_{ij} seja incluída na amostra é dada por:

$$p_{ij} = 1 - \binom{N - m_{ij} - a_{ij}}{n} : \binom{N}{n} \tag{7.2.2.3}$$

O número esperado de unidades secundárias distintas na amostra final é a soma das probabilidades de inclusão, de modo que o tamanho esperado da amostra v, expresso em termos do número equivalente de unidades primárias na amostra final, seja:

$$E(v) = \frac{1}{M} \sum_{i=1}^{N} \sum_{j=1}^{M} p_{ij} \tag{7.2.2.4}$$

Os projetos de amostragem adaptadora de conglomerado levam a estimadores padrões da média populacional e do total populacional tendenciosos. Com as populações espacialmente agregadas, se as unidades adicionais forem acrescentadas à amostra, sempre que se observar grande número de elementos com as características em estudo, a amostra final tenderá a conter valores elevados e maiores que a média populacional, e a média da amostra superestimará a média populacional. Se, por outro lado, o estimador for formado pela média de todos os y valores associados com a seleção de uma unidade primária - as unidades da unidade primária mais as unidades acrescentadas à amostra, como o resultado de uma seleção inicial da unidade primária - a média dessas médias pode subestimar a média populacional, devido ao fato de que toda vez que as unidades com os valores y mais altos que a média forem selecionadas, desta forma a amostragem adicional será feita até que os valores baixos sejam obtidos. Quando unidades com valores baixos são selecionadas, nenhum procedimento compensatório se inicia. Nessa parte, os estimadores serão dados não tendenciosos para os projetos de amostragem adaptadora de conglomerado, desde que estes estimadores sejam de projetos não tendenciosos. A não tendenciosidade não depende de qualquer suposição sobre a população em si.

Uma maneira de se obter um estimador não tendencioso da média da população μ é ignorar todas as unidades acrescentadas adaptativamente à amostra e usar a média da amostra inicial. Isto é, se Y representa o total dos valores y da i'ésima unidade primária ou: $Y_i = \sum_{j=1}^{M} y_{ij}$, o estimador de μ baseado na média da amostra inicial será fornecido por:

$$\bar{y}_0 = \frac{1}{Mn}\sum_{i=1}^{n} Y_i \,. \qquad (7.2.2.5)$$

Este estimador não faz uso das observações acrescentadas adaptativamente à amostra. Isso é de interesse, porque oferece a base para alternativas não adaptadoras, com as quais as estratégias adaptadoras podem ser comparadas. Sabe-se que $\bar{y}_0$ em amostragem aleatória simples é um estimador não tendencioso da média populacional e tem variância:

$$\mathrm{var}(\bar{y}_0) = \frac{N-n}{M^2 Nn} s_0^2 \quad \text{onde} \quad s_0^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - Mm)^2 \,. \qquad (7.2.2.6)$$

Um estimador não tendencioso da variância é:

$$S^2(\bar{y}_0) = \frac{N-n}{M^2 Nn} S_o^2 \quad \text{onde} \quad S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - M\bar{y}_0)^2 \,. \qquad (7.2.2.7)$$

Uma estimativa não tendenciosa da variância não está disponível para amostras sistemáticas com apenas um ponto de partida.

Estimadores não tendenciosos também podem ser obtidos quando se faz o uso das observações adicionais às selecionadas inicialmente. Os estimadores, como o de Hansen-Hurwitz; o estimador de multiplicidade de amostragem de rede e os estimadores usados em amostragem de intercepto de linha, atingem a não tendenciosidade, dividindo-se cada observação por sua probabilidade de seleção e multiplicando-se pelo número de vezes que a unidade foi selecionada. Com os projetos de amostragem adaptadores de conglomerado, entretanto, nem todas as probabilidades de seleção podem ser determinadas a partir dos dados da amostra. Quando uma unidade que não satisfizer a condição aparecer na amostra, poder-se-á não saber se sua probabilidade de seleção foi influenciada pela presença de unidades em sua vizinhança que realmente satisfizeram a condição. Assim, a constante a_{ij}, da qual as probabilidades de seleção dependem, poderá não ser conhecida. O estimador não tendencioso depende apenas dos aspectos das probabilidades de seleção que são conhecidos. Enquanto o estimador de Hansen-Hurwitz atingirá a não tendenciosidade, dividindo-se cada valor de y observado pela probabilidade p_{ij} de seleção, o estimador, efetivamente por $\frac{m_{ij}}{N}$, "parte conhecida" da probabilidade de seleção.

Para a continuidade do desenvolvimento teórico, a seguinte nomenclatura será utilizada:

K número de redes na população, variando de 1 a K

y_k total de valores de y na k'ésima rede.

I_{ik} variável indicadora (I_{ik} =1 se a unidade primária interceptar a k'ésima rede, e I_{ik} =0 caso contrário).

x_k número de unidades primárias da população que interceptam a k'ésima rede.

$$x_k = \sum_{i=1}^{N} \mathrm{I}_{\mathrm{ik}} = m_{ij}$$

A probabilidade de que a unidade primária selecionada intercepte a k'ésima rede é dada por: x_k / N . Para a i'ésima unidade primária, define-se a nova variável w_i como:

$$w_i = \frac{1}{M} \sum_{k=1}^{K} \frac{y_k I_{ik}}{x_k} . \tag{7.2.2.8}$$

O estimador baseado nas probabilidades de seleção parcial é fornecido por:

$$\bar{y}_1 = \frac{1}{n} \sum_{i=1}^{n} w_i . \tag{7.2.2.9}$$

Este estimador é baseado somente nas variáveis para as redes que são interceptadas pelas unidades primárias. Um valor y de rede é utilizado no estimador tantas vezes quantas unidades primárias haja na amostra inicial que a intercepta. As unidades que não satisfazem a condição e não incluídas na amostra inicial, não são utilizadas no estimador. A probabilidade real de seleção para a rede k está relacionada a x_k, mas pode também depender de um modo desconhecido a partir dos dados conhecidos de outras redes.

Desse modo, $\bar{y}_1$ é um estimador não tendencioso de m com variância:

$$Var(\bar{y}_1) = \frac{N-n}{Nn} s_w^2 \qquad \text{onde,} \tag{7.2.2.10}$$

$$s_w^2 = \frac{1}{N-1} \sum_{i=1}^{N} (w_i - m)^2 . \tag{7.2.2.11}$$

Um estimador não tendencioso da variância de y_2 é dado por:

$$S^2(\bar{y}_1) = \frac{N-n}{Nn} S_w^2 \quad \text{, onde} \quad S_w^2 = \frac{1}{n-1} \sum_{i=1}^{n} (w_i - \bar{y}_1)^2 . \tag{7.2.2.12}$$

O estimador de Horvitz-Thompson se torna não tendencioso quando se divide o valor de y, em cada unidade da amostra, pela probabilidade de que a unidade seja incluída na amostra. Nos projetos de amostragem adaptadora de conglomerado nem sempre estas probabilidades de inclusão são conhecidas a partir dos dados da amostra. Em particular, as constantes a_{ij} podem ser desconhecidas desde que a amostra não revele as unidades que não satisfaçam a condição das vizinhanças das unidades da amostra. Desse modo, um estimador não tendencioso é fornecido com base no conhecimento parcial das probabilidades de inclusão que podem ser obtidas dos dados.

Se a_k representa a probabilidade de que uma ou mais unidades primárias, que interceptam a rede k, sejam incluídas na amostra inicial, nos projetos de amostragem adaptadora de conglomerado, esta probabilidade será dada por:

$$a_k = 1 - \binom{N - x_k}{n} : \binom{N}{n} \qquad (7.2.2.13)$$

Se a_{kj} representa a probabilidade de que uma ou mais unidades primárias que interceptam as redes k e j, respectivamente, sejam incluídas na amostra inicial, então para este tipo de projeto tem-se que:

$$a_{kj} = 1 - \left[\binom{N - x_k}{n} + \binom{N - x_j}{n} - \binom{N - x_k - x_j + x_{kj}}{n} \right] : \binom{N}{n} \qquad (7.2.2.14)$$

onde x_k é o número das unidades primárias que intercepta a rede k, e x_{kj} é o número das unidades primárias que interceptam as redes k e j respectivamente.

Os valores dos $a's$ não constituem as probabilidades de inclusão na rede real, mas ao contrário das probabilidades de inclusão, eles são calculados a partir dos dados da amostra.

A variável indicadora z_k assume o valor um (1), quando uma ou mais unidades primárias que interceptam a rede k forem incluídas na amostra inicial, e o valor zero (0), caso contrário. O peso que uma observação recebe do estimador não dependerá, como ocorre com $\overline{y}_1$, do número de unidades primárias interceptadoras selecionadas, enquanto pelo menos uma delas estiver incluída na amostra inicial. Assim, o estimador $\overline{y}_2$ é não tendencioso para a média populacional (m), e sua variância será dada por:

$$Var(\overline{y}_2) = \frac{1}{M^2 N^2} \sum_{k=1}^{K} \sum_{j=1}^{K} y_k y_j \left(\frac{a_{kj}}{a_k a_j} - 1 \right) \tag{7.2.2.15}$$

onde, por convenção, $a_{kk} = a_k$.

Um estimador não tendencioso desta variância, desde que nenhuma das probabilidades conjuntas a_{kj} seja igual a zero (0), será dado por:

$$S^2(\overline{y}_2) = \frac{1}{M^2 N^2} \sum_{k=1}^{K} \sum_{j=1}^{K} \frac{y_k y_j z_k z_j}{a_{kj}} \left(\frac{a_{kj}}{a_k aj} - 1 \right). \tag{7.2.2.16}$$

Nos estudos feitos com o projeto de faixa inicial, as estratégias adaptadoras, com um tamanho de amostra inicial de uma unidade primária, são ligeiramente mais eficientes do que a estratégia não adaptadora. A vantagem relativa das estratégias adaptadoras aumenta com o tamanho da amostra inicial e, também, a eficiência de $\overline{y}_2$ com relação a $\overline{y}_1$ é maior. Com o projeto de amostragem sistemática inicial, as estratégias adaptadoras são muito mais eficientes do que as não adaptadoras. Mesmo com as estratégias sistemáticas convencionais ($\overline{y}_0$ e $\overline{y}_0^*$), as variâncias são menores do que com as estratégias convencionais que usam faixas" (Thompson, 1992)[22].

[22] Tradução livre.

7.2.3 Amostragem Estratificada Adaptadora de Conglomerado

"Na amostragem estratificada adaptadora de conglomerado, uma amostra estratificada inicial é selecionada de uma população. Toda vez que o valor da variável de interesse de qualquer unidade for observado, satisfazendo uma condição especificada, unidades adicionais da vizinhança dessa unidade serão adicionadas à amostra. Outras unidades poderão ser adicionadas à amostra, se qualquer das unidades subseqüentes acrescentadas satisfizerem a condição.

Os projetos de amostragem adaptadora de conglomerado são importantes a partir de um ponto de vista prático, porque existem informações anteriores para muitas populações sobre as quais uma estratificação inicial pode ser baseada, apesar de que a distribuição de padrões de concentração de população pode não ser conhecida. Na amostragem estratificada convencional, as unidades supostamente similares são agrupadas inicialmente em estratos, baseando-se em informação anterior sobre a população, ou simplesmente pela proximidade das unidades. A amostragem adaptadora de conglomerado, por outro lado, fornece um meio de se tirar vantagem de tendências de aglomeração numa população, quando as locações e formas de aglomerados não podem ser previstas antes do levantamento. O projeto de amostragem estratificada adaptadora de conglomerado combina os dois métodos.

Os estimadores convencionais como a média da amostra estratificada são tendenciosos para os projetos adaptadores. Uma complicação que surge na amostragem estratificada adaptadora de conglomerado é que uma seleção em um estrato pode resultar na inclusão de unidades de outros estratos na amostra, de modo que observações em estratos distintos não sejam independentes como na amostragem estratificada convencional. Os estimadores não tendenciosos, que serão apresentados, trabalham com tais cruzamentos de ligações de estratos de maneira diferente.

Para os projetos de amostragem adaptadora de conglomerado, a população é subdividida em L estratos, sendo o estrato H composto de N_h unidades, e o número total de unidades da população denotado por N. Junto com a unidade u_{hi} (a i'ésima unidade do estrato h) está relacionada uma variável de interesse y_{hi}. Para qualquer unidade u_{hi} da população, a vizinhança da unidade u_{hi} é definida como uma coleção de unidades que inclui u_{hi}, e com a probabilidade de que, se a unidade $u_{h'i'}$ estiver na vizinhança da unidade u_{hi}, então a unidade u_{hi} estará na vizinhança da unidade $u_{h'i'}$. A vizinhança da unidade poderá incluir unidades de mais de um estrato. Uma unidade u_{hi} satisfaz a condição de interesse, se o valor de y associado com essa unidade estiver no conjunto C especificado.

Neste projeto, uma amostra inicial de unidades é selecionada de uma população, utilizando-se a amostragem aleatória estratificada. Isto é, dentro do estrato h, uma a.a. simples de n_h unidades é selecionada sem reposição, sendo as seleções (para estratos separados) feitas independentemente. Toda vez que uma unidade selecionada satisfizer a condição, todas as unidades em sua vizinhança, não incluídas na amostra, serão incluídas. Outras unidades poderão ser incluídas na amostra se as unidades adicionadas satisfizerem a condição, de modo que a amostra final contenha toda unidade da vizinhança de qualquer unidade da amostra que satisfaz a condição.

Um exemplo é ilustrado na figura 34, cujo objetivo é estimar a quantidade de uma população conglomerada. Neste caso, estimar as unidades totais por área dos números y de elementos (pontos) dentro de cada unidade. Os elementos (pontos) poderiam representar, por exemplo, a localização de indivíduos com as características em estudo. Uma unidade satisfaz a condição se ela contiver no mínimo um indivíduo com as características, isto é, y≥1. A população foi dividida em cinco estratos e uma a.a. simples, de três unidades cada uma, foi selecionada dentro de cada estrato. A amostra simples final está representada na figura 35. A vizinhança de uma unidade consiste nessa mesma unidade mais as unidades adjacentes nos sentidos norte-sul e leste-oeste. Ao se aplicar o projeto de amostragem estratificada adaptadora de conglomerado, tem-se a amostra final, que está representada pela figura 35.

A população pode ser subdividida em k conjuntos de unidades (redes) tais que a seleção da amostra inicial, de qualquer unidade numa rede, resultará em inclusão na amostra final de todas as unidades dessa rede. Uma unidade que não satisfaça a condição, consiste de uma rede formada apenas de si mesma. A seleção inicial de uma unidade que satisfaça a condição, resultará na adição a uma amostra, não somente de todas as unidades dessa rede, mas, também, de unidades fora da rede, isto é, unidades que não satisfaçam a condição senão na vizinhança de um ou mais membros da rede. Na figura 36, as redes interceptadas pela amostra inicial estão esquematizadas por linhas pretas mais largas. As outras unidades da amostra (unidades limites) não satisfazem a condição e não pertenciam à amostra inicial. Mas cada uma pertence à vizinhança de uma ou mais unidades que satisfazem a condição das redes que interceptam a amostra inicial. O número de vezes que uma unidade é selecionada iguala ao número de unidades de sua rede, ou de uma rede que intercepta sua vizinhança, selecionada na amostra inicial.

Se r_{hi} representar o número de vezes que a unidade u_{hi} foi selecionada e m_{khi},o número de unidades da interseção do estrato k com a rede que contenha a unidade u_{hi}, e para uma unidade u_{hi} que não satisfaça a condição, se k.h.i representar o número total de unidades na interseção do estrato k com a coleção de redes distintas, com exclusão de u_{hi}. A seleção inicial de qualquer dessas a_{khi} unidades resultará na adição da unidade u_{hi} na amostra. Se a_{khi} assumir o valor zero (0) para qualquer unidade u_{hj} que satisfaça a condição, então o número esperado de vezes que a unidade u_{hi} é selecionada, é dado por:

$$E(r_{hi}) = \sum_{k=1}^{L} n_k \frac{m_{khi} + a_{khi}}{N_k} \tag{7.2.3.1}$$

FIGURA 34: Amostragem por Conglomerado

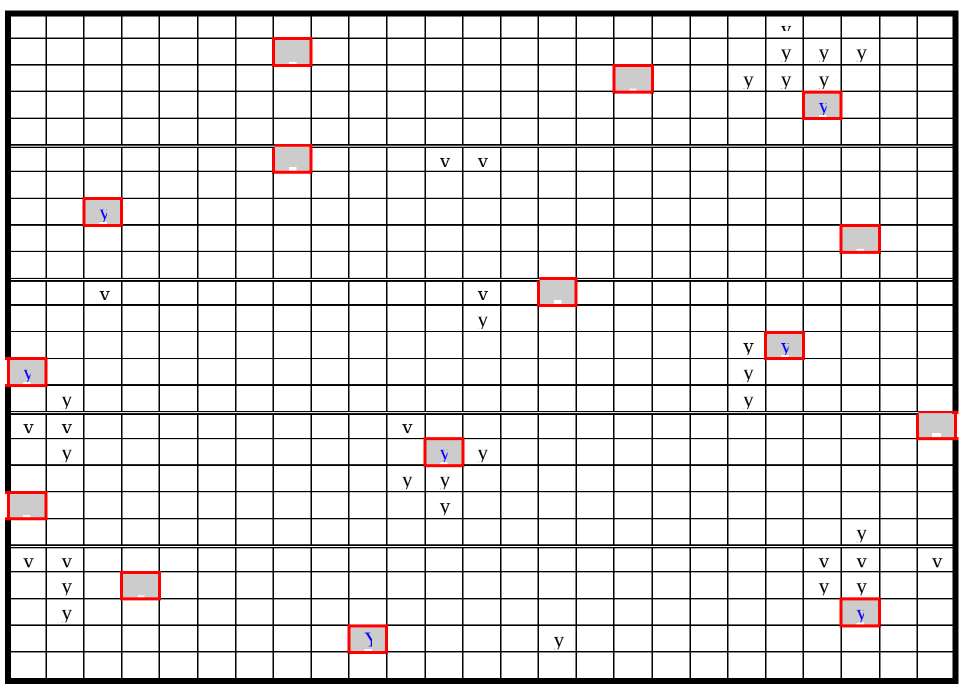


FIGURA 35: Amostragem Sistemática Adaptadora de Conglomerado

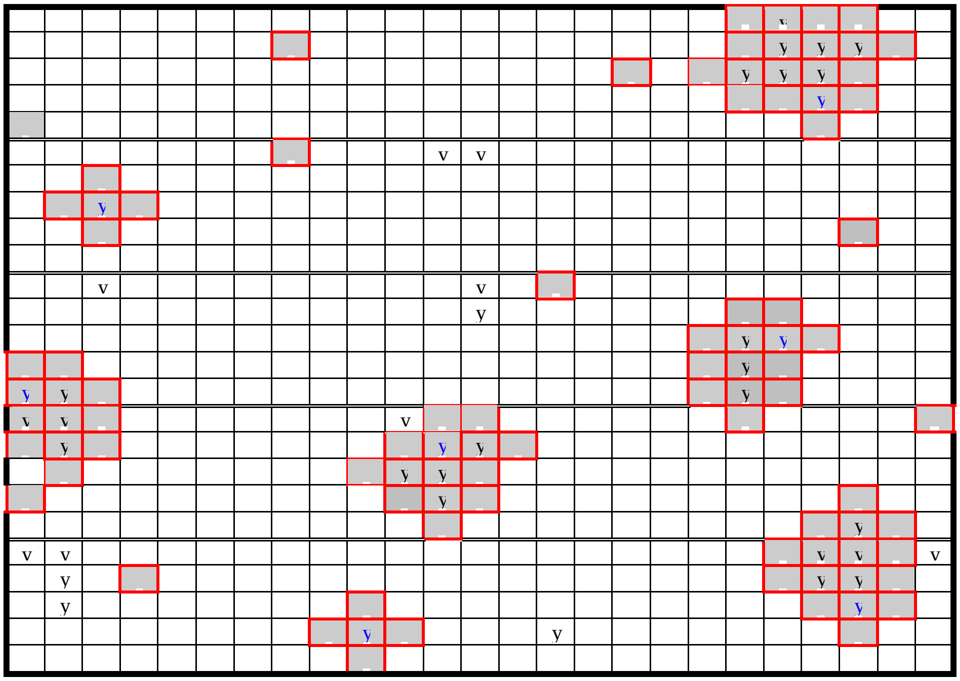
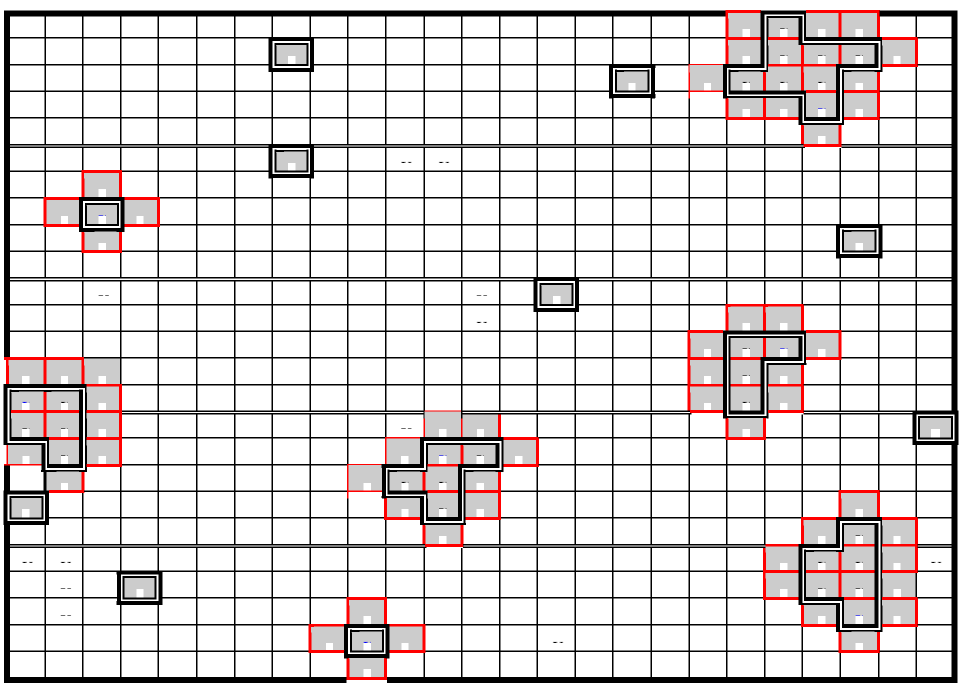


FIGURA 36: Redes Interceptadas Pela Amostra Inicial



A unidade u_{hi} será incluída na amostra se uma ou mais unidades da rede a que u_{hi} pertença, estiverem incluídas na seleção inicial ou se uma unidade u_{hi}, que não satisfizer a condição, estiver incluída na amostra inicial. Por causa da amostragem aleatória estratificada inicial, a probabilidade de inclusão p_{hi} para a unidade u_{hi} será dada por:

$$p_{hi} = 1 - \prod_{k=1}^{L} \binom{N_k - m_{khi} - a_{khi}}{n_k} \div \binom{N_k}{n_k}. \tag{7.2.3.2}$$

O número de elementos (v) esperado para a amostra, isto é, o número esperado de unidades distintas da amostra final é a soma das N probabilidades de inclusão da população.

Os estimadores convencionais, como a média da amostra estratificada, embora não tendenciosos para a média da população, quando se faz o uso da amostragem aleatória estratificada clássica, são tendenciosos para os projetos adaptadores. Um estimador não tendencioso da média da população ($\overline{y}_0$) pode ser obtido, entretanto, utilizando-se o estimador estratificado convencional para a média, baseado na amostra inicial (ignorando-se todas as observações subseqüentes).

Para amostragem (com reposição) de probabilidades de seleção conhecidas, o estimador de Hansen-Hurwitz atinge a não tendenciosidade quando se divide o valor (y) de cada unidade pela probabilidade de seleção retirada por retirada dessa unidade. Mais precisamente, cada observação é dividida pelo número previsto de vezes que ela é selecionada na amostra e multiplicada pelo número de vezes que ela foi observada.

Com a amostragem adaptadora estratificada de conglomerado, as probabilidades de seleção e o número esperado de vezes não são conhecidos para todas as unidades selecionadas da amostra, de modo que um estimador não tendencioso deve basear-se somente nos aspectos dos números da seleção esperada, que podem ser determinados a partir dos dados.

Para a unidade u_{hi}, define-se a nova variável w_{hi} (total de valores y da rede à qual u_{hi} pertence) cujo peso é a fração de amostragem do estrato, e se divide o valor pela soma ponderada dos tamanhos da interseção do estrato da rede como se segue:

$$w_{hi} = \frac{n_h}{N_h} \sum_{k=1}^{L} e_{khi} \Big/ \sum_{k=1}^{L} \frac{n_k}{N_k} m_{khi} \qquad (7.2.3.3)$$

onde e_{khi} é o total de valores y da interseção do estrato k com a rede que inclui a unidade u_{hi}, e m_{khi} é o número de unidades desta interseção. Assim, um estimador para a média populacional é dado por:

$$\overline{\hat{y}}_1 = \frac{1}{N} \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} w_{hi} \ . \qquad (7.2.3.4)$$

Se r_{khi} é a variável aleatória que representa o número de unidades da amostra inicial, que pertencem à interseção do estrato k e a rede a qual pertence a unidade hi, o estimador pode ser definido da seguinte forma:

$$\overline{\hat{y}}_1 = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \left[y_{hi} \sum_{k=1}^{L} r_{khi} \Big/ \left(\sum_{k=1}^{L} \frac{n_k}{N_k} m_{khi} \right) \right]. \qquad (7.2.3.5)$$

Uma vez que $E(r_{khi}) = n_k m_{khi} / N_k$ segue-se que $\overline{\hat{y}}_1$ é um estimador não tendencioso da média da população.

Como w_{hi} representa a variável de interesse para a variável u_{hi}, para cada unidade da população, $\overline{\hat{y}}_1$ é a média de uma amostra aleatória estratificada, portanto tem variância dada por:

$$Var(\overline{\hat{y}}_1) = \frac{1}{\mathrm{N}^2} \sum_{\mathrm{h}=1}^{\mathrm{L}} \mathrm{N_h} (N_h - n_h) \frac{s_h^2}{n_h} \qquad (7.2.3.6)$$

A variância e a média da população do estrato são respectivamente:

$$s_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{n_h} (w_{hi} - \hat{W}_h)^2 \text{ e } \overline{W_h} = \frac{1}{N_h} \sum w_{hi} \; . \qquad (7.2.3.7)$$

Assim, um estimador não tendencioso para a variância da média da amostra será fornecido por:

$$S_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (w_{hi} - \hat{w}_h)^2 \qquad (7.2.3.8)$$

usando-se a média da amostra $\overline{w}_h = (1/n_h) \sum w_{hi}$.

Uma variação do estimador $\hat{\overline{y}}_1$ pode ser construída através da relação com o estimador estratificado de multiplicidade de amostragem de rede, em que o peso que uma observação recebe depende do estrato em que a amostra inicial intercepte a rede dessa unidade. Para a unidade u_{hi} define-se a nova variável w'_{hi} como sendo o total dos valores de y de toda a rede à qual a unidade hi pertença, dividido pelo número total de unidades dessa rede:

$$w'_{hi} = \sum_{k=1}^{L} e_{khi} / \sum_{k=1}^{L} m_{khi} \; . \qquad (7.2.3.9)$$

O estimador de multiplicidade estratificado é dado, substituindo-se w_{hi} por w'_{hi} no estimador da média populacional:

$$\hat{\overline{y}}_1{}' = \frac{1}{N} \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} w'_{hi} \; . \qquad (7.2.3.10)$$

Toda vez que qualquer unidade de uma rede for selecionada na amostra inicial, o estimador inclui um termo com o total dos valores de y, para esta rede, dividido pelo tamanho da rede e ponderado por N_k / n_k para o estrato do qual a unidade foi selecionada. Assim, cada unidade individual de y aparece no estimador toda vez que qualquer unidade de rede, à qual ele pertença, for selecionada na amostra inicial, mas com pesos que dependem dos estratos dos quais vieram as seleções iniciais. Desta forma, o estimador $\hat{\bar{y}}_1'$ pode ser escrito na forma alternativa:

$$\hat{\bar{y}}_1' = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(y_{hi}\sum_{k=1}^{L}\frac{N_k}{n_k} r_{khi} / \sum_{k=1}^{L} m_{khi} \right) \tag{7.2.3.11}$$

A variância estimada de $\hat{\bar{y}}_1'$ é fornecida pelas equações dadas para $\hat{\bar{y}}_1$, em que basta substituir w por w'. Também é possível usar um estimador que ignore todas as unidades adicionadas nos limites do estrato interceptador. Para este estimador, usa-se a nomenclatura w_{hi}'' para representar o total de valores de y que estejam na interseção do estrato com a rede de unidade u_{hi} dividido pelo número de unidades desta interseção. O estimador e sua variância são os mesmos anteriores, com a diferença da utilização de w'' no lugar de w. A não tendenciosidade e outras propriedades seguem o caso não estratificado, desde que as componentes dos estratos sejam independentes.

Para cada projeto em que sejam conhecidas as probabilidades de inclusão, o estimador de Horwitz-Thompson será não tendencioso quando o valor de y para cada unidade da amostra for dividido pela probabilidade de que a unidade seja incluída na amostra. Com a amostragem adaptadora de conglomerado, estas probabilidades de inclusão não podem ser determinadas a partir dos dados da amostra. Entretanto, um estimador pode ser construído, usando-se para cada unidade a probabilidade de que a amostra inicial intercepte a rede à qual esta unidade pertença, e dando peso zero (0) para qualquer observação que não satisfaça a condição da amostra inicial.

Se as K redes distintas forem numeradas de 1 a K, sem considerar os limites do estrato, e y_i representar o total dos valores de y na i'ésima rede da população, sendo x_{hi} o número de unidades no estrato h, que intercepta a rede i, então a probabilidade α_i de que a amostra inicial intercepte a rede i é dada por:

$$a_i = 1 - \prod_{k=1}^{L} \binom{N_k - x_{ki}}{n_k} \div \binom{N_k}{n_k}. \tag{7.2.3.12}$$

Se $q_i = 1 - a_i$, então a probabilidade a_{ij} de que a amostra inicial intercepte as redes i e j respectivamente, será dada por:

$$a_{ij} = 1 - q_i - q_j + \prod_{k=1}^{L} \binom{N_k - x_{ki} - x_{kj}}{n_k} \div \binom{N_k}{n_k}. \tag{7.2.3.13}$$

Se a variável indicadora z_i assumir o valor um (1), quando a mostra inicial interceptar a rede i, e zero (0) caso contrário, o estimador estratificado do tipo Horwitz-Thompson modificado será dado por:

$$\hat{y}_2 = \frac{1}{N} \sum_{i=1}^{K} \frac{y_i z_i}{a_i} \ . \tag{7.2.3.14}$$

Para i=1, 2, ... , K, a variável z_i é uma v.a. de Bernoulli com $E(z_i) = a_i$, $Var(z_i) = a_i(1 - a_i)$, e $Cov(z_i, z_j) = a_{ij} - a_i . a_j$, para $i \neq j$. Então $\hat{y}_2$ é um estimador não tendencioso da média populacional, e a variância do estimador será dada por:

$$Var(\hat{y}_2) = \frac{1}{N^2} \sum_{i=1}^{K} \sum_{j=1}^{K} y_i y_j \left(\frac{a_{ij}}{a_i . a_j} - 1 \right) \tag{7.2.3.15}$$

onde, por convenção, $a_i = a_{ii}$.

Um estimador não tendencioso para esta variância, desde que $E(z_i z_j) = a_{ij}$, é dado por:

$$S^2(\hat{\bar{y}}_2) = \frac{1}{N^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{y_i y_j z_i z_j}{a_{ij}}\left(\frac{a_{ij}}{a_i a_j} - 1\right), \qquad (7.2.3.16)$$

uma vez que a probabilidade conjunta a_{ij} da interseção seja diferente de zero (0) para qualquer par de redes.

O estimador $\hat{\bar{y}}_2$ não é o verdadeiro estimador de Horwitz-Thompson, uma vez que as probabilidades de interseção a_i não são idênticas às probabilidades de inclusão para o projeto de amostragem adaptadora de conglomerado. O tamanho previsto para a amostra e outras propriedades da estratégia de amostragem dependem das probabilidades reais de inclusão, que foram colocadas anteriormente.

Existe uma tendência para a vantagem relativa das estratégias adaptadoras de amostragem ao serem comparadas com as estratégias não adaptadoras, isto é, a eficiência da estratégia adaptadora é maior do que a não adaptadora.

A amostragem estratificada convencional tem variância maior que a estratégia estratificada adaptadora de conglomerado" (Thompson, 1992)[23].

[23] Tradução livre.

ANEXO 08

GLOSSÁRIO

GLOSSÁRIO

- **Aleatório**: Que depende de acontecimentos incertos, favoráveis ou não a um determinado evento.

- **Amostra**: Pequena parte ou porção representativa de uma população.

- **Amostragem**: Técnica de pesquisa na qual um sistema preestabelecido de amostras é considerado idôneo para representar a população ou o universo pesquisado, com margem de erro aceitável.

- **Atributo**: Aquilo que é próprio ou peculiar de alguém ou alguma coisa. Condição, propriedade, qualidade.

- **Campo**: Domínio sobre o qual se exerce uma atividade, um estudo.

- **Características**: Conjunto de qualidades próprias que descrevem os elementos de uma população ou de uma amostra.

- **Centro Mediano**: Para dados espacializados a mediana é representada pelo centro mediano, que é definido pela interseção de duas retas perpendiculares entre si, geralmente Norte/Sul e Leste/Oeste, que divide a distribuição dos dados em dois conjuntos com o mesmo percentual de elementos, 50% em cada um (Gerardi & Silva, 1981).

- **Centro Médio**: Em Estatística espacial o centro médio é similar à média univariada, e é definido como o ponto de um plano que minimiza a soma das distâncias quadráticas a todos os outros pontos do plano, ou ainda como o ponto de equilíbrio de um dado plano na ponta de um prisma. Calcula-se a média aritmética da variável X e depois da variável Y, assim, o centro médio será dado pelo ponto: $(\bar{x}, \bar{y})$ (Gerardi & Silva, 1981)..

- **Centro Modal**: O centro modal é representado pela porção do espaço que tem maior intensidade de ocorrência de um dado fenômeno (Gerardi & Silva, 1981).

- **Coletar**: Ato de retirar de uma população elementos.

- **Conglomerado**: Estrutura cujos elementos, unidos, representam uma unidade; união de partes num todo.

- **Condição**: Algo preestabelecido como requisito para que outra coisa seja feita ou entre em vigor. Obrigação que se impõe ou se aceita como parte essencial de uma teoria.

- **Contínuo**: Ininterrupto, que não tem suas partes separadas. Espaço infinito pertencente ao domínio dos números reais.

- **Correlação**: Interdependência entre variáveis. Ligação entre duas características (correlação simples) ou mais (correlação múltipla), tal que as variações dos seus valores tenham sempre o mesmo sentido (correlação positiva) ou sentidos opostos (correlação negativa). O coeficiente de correlação linear representa o quociente entre a covariância de duas variáveis pelo produto de seus desvios quadráticos médios. (Este coeficiente é tão mais próximo de +1 ou de -1 quão mais estreita for a ligação positiva ou negativa entre as variáveis).

- **Correlograma**: Matriz com todas as correlações entre as variáveis.

- **Cota**: Valor que representa o número de elementos com determinada(s) característica(s) a serem amostrados.

- **Covariância**: Covariância entre duas variáveis aleatórias X e Y é a esperança matemática do produto das variáveis centradas associadas a X e a Y, isto é, $Cov(X,Y) = E[(X - m_x)(Y - m_y)]$.

- **Desvio-padrão**: Medida de variabilidade em torno da média; raiz quadrada da variância.

- **Detecção**: Ato ou efeito de detectar.

- **Discreto**: Por pontos ou sinais separados. Espaço infinito ou finito enumerável.

- **Distância Padrão**: Representa a variabilidade de um conjunto de pontos em torno de um valor médio central, motivo pelo qual a expressão gráfica da distância padrão resulta num círculo centrado no centro médio, cujo raio é a medida da distância padrão. A fórmula de cálculo é: $d = \sqrt{(\sum_{i=1}^{n}(x_i - \overline{x})^2 . \sum_{i=1}^{n}(Y_i - \overline{y})^2)} / \sqrt{n-1}$.

- **Espaço Amostral**: É o conjunto de todos os resultados possíveis de um experimento. Quando o espaço amostral consiste em um número infinito ou finito numerável de eventos, é chamado de espaço amostral discreto; se consiste de todos os números reais de determinado intervalo, é um espaço amostral contínuo.

- **Experimento Aleatório**: É o processo de coleta de dados relativos a um fenômeno que acusa variabilidade em seus resultados.

- **Estratégia**: Procedimento que define os objetivos a alcançar, e a ciência que determina a maneira de atingi-los.

- **Estatística**: Ciência que tem por objetivo a coleção, análise e interpretação de dados numéricos a respeito de fenômenos coletivos ou de massa, bem como a indução das leis a que tais fenômenos obedecem e, ainda, a representação numérica e comparativa, em tabelas ou gráficos, dos resultados da análise desses fenômenos. Os fatos ou resultados numéricos pertencentes a um fenômeno coletivo ou de massa são representados por meio das estatísticas.

- **Estimação**: Procedimento para se estimar um valor populacional (parâmetro), isto é, encontrar um valor aproximado do parâmetro do qual depende uma lei de distribuição estatística ou uma de suas características, baseando-se em uma informação fornecida por uma amostra. Em estatística, a estimação pode ser pontual ou intervalar.

- **Estimador**: Que, ou o que se estima.

- **Estimador de máximo-verossimilhança**: Estimador fornecido pela função de máximo-verossimilhança dada por: $L(q) = \prod_{i=1}^{n} f(x,q)$, onde q representa o parâmetro a ser estimado através de uma a.a. de n elementos.

- **Estimativa**: Resultado obtido por meio de uma apreciação ou de uma avaliação (estimação). Operação que consiste em determinar, em termos probabilísticos, a partir dos resultados observados em uma ou mais amostras de uma população, os parâmetros dessa população ou a lei de probabilidade que se quer representar.

- **Estrato**: Cada um dos conjuntos em que se divide uma população da qual se deve tirar uma amostra.

- **Estratégia**: Arte de dirigir um conjunto de disposições: estratégia política.

- **Evento**: Acontecimento, sucesso. Em um fenômeno aleatório, ocorrência de um elemento de um conjunto universo definido previamente; acontecimento.

- **Feição**: Categoria

- **Fenômeno**: Fato natural constatado, suscetível de estudo científico, e que pode tornar-se objeto de experiência.

- **Finito**: Limitado. Conjunto cujo número de elementos pode ser expresso por um número natural. (Dedekind estabeleceu que um conjunto é finito se é impossível estabelecer uma bijeção entre eles e uma de suas parte próprias).

- **Independência**: Autonomia. Dois eventos A e B de um espaço de resultados E se dizem independentes $\Leftrightarrow P(A \cap B) = P(A).P(B)$.

- **Índice**: Número que permite resumir e caracterizar a variação relativa de uma grandeza simples ou complexa entre duas situações, das quais uma serve de base.

- **Indivíduo**: Unidade pertencente a uma população ou a uma amostra.

- **Inferência Estatística**: Conjunto dos métodos que permitem formular, em termos probabilísticos, um julgamento sobre uma população a partir dos resultados observados em uma amostra extraída ao acaso dessa população.

- **Inspeção**: Ação de ver, olhar e observar.

- **Média**: Quantidade que ocupa lugar eqüidistante entre dois extremos; quociente da soma dos valores de uma característica quantitativa pela quantidade total de valores.

- **Mediana**: A mediana divide o grupo de dados (em ordem crescente) em dois grupos, isto é, a mediana é o valor que deixa nele e ou abaixo dele 50% dos menores valores, e nele e ou acima dele 50% dos maiores valores.

- **Nível de Confiabilidade**: Representa a probabilidade de que o valor real do parâmetro populacional esteja entre os limites do intervalo gerado através da distribuição de probabilidades da estatística amostral.

- **Nível de significância**: Representa a probabilidade máxima de se rejeitar uma hipótese a respeito do parâmetro populacional, sendo a mesma verdadeira.

- **Paradigma**: Modelo, padrão, norma; exemplo. Na lingüística estrutural, conjunto das unidades que são substituíveis ou comutáveis entre si num determinado contexto.

- **Parâmetros**: Grandeza mensurável que permite apresentar de maneira mais simples e mais geral as principais características de uma população.

- **População**: Conjunto de indivíduos (seres humanos, seres vivos, objetos inanimados) submetidos a um estudo estatístico. (Universo).

- **Precisão**: Qualidade de um mecanismo bem regulado, exato, justo, e que, portanto, pode desempenhar satisfatoriamente sua função. Qualidade de uma medida ou de um instrumento capaz de fornecer com grande aproximação o mesmo resultado, quando a medição é repetida por inúmeras vezes. (A média de uma amostra está bem próxima da média populacional ou do valor esperado).

- **Previsão**: O que se prevê.

- **Primário**: Que precede outro no tempo e no espaço; primeiro.

- **Probabilidade**: Qualidade de uma afirmação que tem as maiores condições de estar em conformidade com a realidade, sem que se possa demonstrar absolutamente a verdade. A probabilidade de um evento é representada pela razão entre o número de casos favoráveis pelo número de casos possíveis de um evento qualquer, em que os casos possíveis têm supostamente o mesmo grau de possibilidade; na teoria axiomática.

- **Projeto de Amostragem**: Plano de realizar coleta de elementos que representem a população.

- **Quota**: Cota.

- **Rede**: Malha de pontos que têm características segundo especificações.

- **Referenciar**: Tomar como ponto de referência.

- **Região**: Grande extensão territorial; território caracterizado por condições especiais de clima, aspecto físico, população, produção, posição geográfica, etc. Em Geografia, a definição de região foi elaborada no seio da Geografia Possibilista Francesa no início do século XX, o conceito de região difundiu-se até constituir, na atualidade, o centro de interesse de um campo específico desta disciplina: a Geografia Regional. Inicialmente, a região era definida como uma porção homogênea, devido à paisagem, com fisionomia própria. Gradativamente, a definição de região foi sendo estendida como uma realidade econômica e social.

- **Secundário**: Que é de segunda ordem.

- **Seleção**: Escolha feita a partir de critérios e objetivos bem definidos.

- **Sofisticação Política**: o conceito utilizado no texto refere-se à elaboração de Neuman W. Russel, em The Paradox of Mass Politics. Knowledge and Opinion in the American Electorate. O autor propõe três dimensões para mensurar a sofisticação Política do cidadão, adquirida através de um processo de conhecimento crescente: 1 - saliência política, entendida como interesse, preocupação e atenção em relação à política; 2 - conhecimento político, que é a familiaridade com os principais issues (questões) o debate político, figuras públicas e eventos políticos importantes; 3 - conceitualização da política, que está relacionada à organização cognitiva por meio de conceitos abstratos e o uso ativo do conhecimento político na avaliação das principais questões (issues) que estão colocadas no debate político.

- **Tendencioso**: O que revela tendência (propensão) ao representar a realidade.

- **Teste de Hipóteses**: Método estatístico para verificar valores de parâmetros populacionais. O teste de hipóteses está diretamente relacionado com a distribuição de probabilidades das estatísticas amostrais.

- **Transversal**: Que passa ou está na posição oblíqua.

- **Variância**: Medida de variabilidade em torno do valor central (média). A variância para uma amostra aleatória simples é dada por: $\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$.

- **Variável**: Quantidade que pode tomar sucessivamente diferentes valores no decurso de um mesmo cálculo.

- **Variável aleatória**: Regra de associação de um valor numérico a cada ponto do espaço amostral.