



UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
Instituto de Geociências e Ciências Exatas
Campus de Rio Claro

Estudo dos Planos Amostrais e Estimadores para a Aplicação no Dimensionamento da População Canina de Rio Claro - SP

Olívio Crispim de Medeiros

Dissertação apresentada ao Programa de Pós-Graduação – Mestrado Profissional em Matemática Universitária do Departamento de Matemática como requisito parcial para a obtenção do grau de Mestre

Orientadora
Profa. Dra. Selene Maria Coelho Loibel

2013

310 Medeiros, Olívio Crispim
M488e Estudo dos Planos Amostrais e Estimadores para a Aplicação no Dimensionamento da População Canina de Rio Claro - SP/ Olívio Crispim de Medeiros- Rio Claro: [s.n.], 2013.
87 f. : fig., tab.

Dissertação (mestrado) - Universidade Estadual Paulista, Instituto de Geociências e Ciências Exatas.
Orientadora: Selene Maria Coelho Loibel

1. Estatística. 2. Técnicas de Amostragem. I. Título

Ficha Catalográfica elaborada pela STATI - Biblioteca da UNESP
Campus de Rio Claro/SP

TERMO DE APROVAÇÃO

Olívio Crispim de Medeiros

ESTUDO DOS PLANOS AMOSTRAIS E ESTIMADORES PARA A
APLICAÇÃO NO DIMENSIONAMENTO DA POPULAÇÃO CANINA DE
RIO CLARO - SP

Dissertação APROVADA como requisito parcial para a obtenção do grau de Mestre no Curso de Pós-Graduação em Matemática Universitária do Instituto de Geociências e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, pela seguinte banca examinadora:

Profa. Dra. Selene Maria Coelho Loibel
Orientadora

Prof. Dr. José Sílvio Govone
DMAC - Unesp-RC

Profa. Dra. Ermínia de Lourdes Campello Fanti
IBILCE - Unesp-RP

Rio Claro, 21 de Outubro de 2013

À MEUS PAIS (in memorian), pelo dom da vida, pela força e carinho que sempre dedicaram para a realização de minhas atividades, que me ensinaram a amar de verdade e que foram meus primeiros mestres nesta brilhante caminhada.

Agradecimentos

A DEUS, o Grande arquiteto do Universo, pelo dom da sabedoria e inteligência, pelo saber que construímos, pelo que somos e pela nossa existência.

A minha esposa e Filhos que tanto incentivaram a conclusão deste curso e orgulham-se do meu trabalho.

Às Prof. e Coordenadoras Alice Kimie Miwa Libardi e Suzinei Aparecida Siqueira Marconato com quem mantive meu primeiro contato no programa e nesta universidade, pelas suas grandiosas contribuições.

À Prof. Dr^a Selene Maria Coelho Loibel, pela orientação, amizade e exemplo, pela paciência e esforços dedicados e confiança no meu trabalho, o qual veio me proporcionar um grande crescimento acadêmico.

Aos Prof. Dr. Vladimir Seixas e Marta Cilene Gadotti pelos seus exemplos de dedicação e amor às suas profissões. O prof. Vlademir além de ter sido meu primeiro professor neste programa foi também membro ativo na banca de minha qualificação.

Ao Prof. Dr. José Sílvio Govone e a Profa. Dra. Ermínia de Lourdes Campello Fanti, por suas valiosas contribuições para o texto final.

A todos os Professores do Departamento de Matemática da Unesp, do programa "Mestrado Profissional" e demais funcionários que participaram de forma direta ou indireta com suas colaborações.

À Maria Nazareth da Nóbrega Oliveira pelo apoio e incentivo aos meus estudos, especialmente na educação básica.

A Universidade Estadual do Maranhão, especialmente ao CESBA, pela oportunidade da graduação, do meu trabalho e incentivo para a conclusão deste curso.

Aos irmãos da loja Maçônica "Amizade Fraternal 275" pela recepção e acolhida colorosa não só em sua loja como também nesta cidade.

Aos colegas e amigos do CEM - Pe. Fábio Beragnolli pela compreensão, apoio e incentivo desta árdua e gratificante caminhada.

A todos os colegas de minha turma, Antônio Nilson, Carlos, Denis, Edgar, Edmar, João Paulo, Juliano, Leandro, Mariana, Renato e Vinícius, que juntos enfrentamos todos os desafios que nos foram impostos.

À meus irmãos, Bento, Bonifácio, Fátima e Leonardo, pelo incentivo constante na busca da ascensão de mais um degrau nesta escada do conhecimento.

À Aline Fátima Weber do CEM - Pe. Fábio Bertagnolli e Suzan Cleyde Martins Figueirêdo da UEMA, pelas correções ortográficas e do inglês, respectivamente.

*"Vivei na paz austera dos laboratórios e bibliotecas.
Dizei, dentro de vós, a princípio: - que fiz eu pela minha instrução?
Depois, à medida que fordes progredindo: - que fiz eu pela Pátria?
Quem sabe se não tereis, algum dia, a ventura de pensar que, de algum modo,
contribuistes para o bem da humanidade?
Pode a vida oferecer-vos, mais ou menos, tais oportunidades. Mas é necessário,
quando se chega ao fim, ter o direito de dizer: - fiz o que pude".*

Pasteur

Resumo

O conhecimento dos métodos de amostragem juntamente com suas técnicas para a coleta de dados é indispensável a todos àqueles que desejam realizar trabalhos com levantamentos amostrais. Inicia-se o seu estudo fazendo-se um apanhado de fatos históricos que marcaram a estatística e a amostragem juntamente com seus principais pensadores desde o século XVIII. Em seguida são descritos os principais estimadores da amostragem aleatória simples (*AAS*) com ou sem reposição onde as mesmas são comparadas. Tem-se também a amostragem estratificada (*AE*), destacando-se as alocações Proporcional, Uniforme e Ótima de Neyman com seus respectivos estimadores, feita uma comparação entre *AAS* e *AE* para uma indicação da melhor entre as técnicas descritas. Outras técnicas também são mencionadas de forma superficial. Um exemplo de aplicação é apresentado no final do capítulo 3 para uma visualização numérica das teorias apresentadas. Por fim, tem-se alguns resultados de um levantamento de dados com aplicação direta ao projeto de extensão "Dimensionamento da População Canina no município de Rio Claro - SP".

Palavras-chave: Estatística, Técnicas de Amostragem.

Abstract

The knowledge of sampling methods along with their techniques for data collection is indispensable to all those wishing to undertake work with sample surveys. We begin this study with an overview of historical events that marked statistics and sampling along with its major thinkers since the eighteenth century. Then we describe the main estimate of simple random sampling (SRS), with or without replacement, where the same are compared. We also present stratified sampling (SS), highlighting Neyman's Proportional, Uniform and Optimal allocations with their respective estimate. A comparison is made between the SRS and SS for an indication of the best among the techniques described. Other techniques are also mentioned, in a superficial way. An application example is presented at the end of chapter 3 for a numerical display of the theories presented. Finally, there are some results of a survey of data with direct application to the extension project "Scaling of the Canine Population in the city of Rio Claro - SP".

Keywords: Statistics, Sampling Techniques.

Lista de Figuras

4.1	% do nº de domicílios com 1, 2, 3, 4, 5 ou mais cães	68
4.2	Distribuição dos cães por categoria OMS - %	69
4.3	% de animais vacinados contra raiva por estrato	70
4.4	Controle de Natalidade - %	71
4.5	Controle de Natalidade das Fêmeas - %	71
4.6	% de domicílios cujas pessoas dão assistência a 1, 2, 3, 4, 5 ou mais cães	75
4.7	Percentual dos cães de raça	76

Lista de Tabelas

3.1	Valores dos Y_i (Áreas Cultivadas)	58
3.2	Distribuição dos elementos da amostra por estrato	59
3.3	Resumo dos cálculos por estrato	60
4.1	Alocação ótima de Neyman	66
4.2	Distribuição do N° de domicílios por estratos e entrevistados	67
4.3	Percentual do n° de domicílios com o n° de cães que possuem	67
4.4	Porcentagem dos cães segundo a categoria da OMS a que pertencem	68
4.5	Total de cães distribuídos por estrato e sexo	69
4.6	Total de cães distribuídos por estrato, categoria e sexo	70
4.7	Cães da categoria 1, Vacinação, por estrato e sexo	72
4.8	Cães da categoria 1, Machos por estrato, Castração	72
4.9	Cães da categoria 1, Fêmeas por estrato, Controle de Natalidade	73
4.10	Cães da categoria 2, critério de Vacinação, por estrato e sexo	73
4.11	Cães da categoria 2, Machos por estrato, Castração	74
4.12	Cães da categoria 2, Fêmeas por estrato, Controle de natalidade	74
4.13	Cães de vizinhança que recebem assistência	75
4.14	Resumo dos cálculos por estratos	76
A.1	Área cultivada com milho, em ha, de 194 propriedades de Lavras - MG, na safra de 1988	83
B.1	Percentuais da População Canina - Machos	86
B.2	Percentuais da População Canina - Fêmeas	87

Sumário

1	Introdução	19
2	Fundamentos	21
2.1	Revisão Bibliográfica	21
2.2	Objetivos	22
2.3	Panorama Histórico	22
2.3.1	Da Estatística	22
2.3.2	Da Amostragem	24
3	Técnicas de Amostragem	27
3.1	Definições	28
3.2	Amostra Aleatória Simples - AAS	30
3.2.1	Amostragem Aleatória Simples com Reposição - AAS_C	30
3.2.2	Amostra Aleatória Simples sem Reposição - AAS_S	36
3.2.3	Comparação entre AAS_C e AAS_S	41
3.3	Amostragem Estratificada - AE	41
3.3.1	Notação e Relações Úteis	42
3.3.2	Métodos de Estratificação	45
3.3.3	Definição do plano amostral nos estratos e alocação	47
3.3.4	Efeito do Planejamento	51
3.3.5	Estimação das Proporções	53
3.3.6	Intervalos de Confiança	55
3.3.7	Determinação do Tamanho da Amostra	55
3.4	Amostragem por Conglomerados - AC	56
3.5	Amostragem em Múltiplos Estágios - AME	57
3.6	Amostragem Sistemática - AS	57
3.7	Exemplo de Aplicação	58
4	Aplicação com dados reais	61
4.1	O Projeto: População Canina	61
4.1.1	Introdução	61
4.1.2	A técnica Pasteur São Paulo	63

4.1.3	Entrevistas domiciliares	63
4.1.4	Contagem dos cães fora dos domicílios	63
4.2	Resultados e Discussões	64
4.2.1	Campo da Pesquisa	64
4.2.2	Análise Estatística	66
4.2.3	Estimação das Proporções	78
5	Conclusão	79
	Referências	81
A	Dados de Áreas cultivadas com Milho em Lavras - MG	83
A.1	Dados para estimação de uma amostra aleatória simples e estratificada	83
B	Dados Gerais da População Canina	85

1 Introdução

Através do presente trabalho serão estudadas técnicas de amostragem utilizadas em pesquisas de levantamento de dados com seus respectivos estimadores com o objetivo de propor àquele método que melhor se adapte, no que se refere a apresentação dos melhores resultados, para a aplicação no dimensionamento e classificação da população canina na cidade de Rio Claro - SP.

É sabido que o uso inadequado de um procedimento amostral pode levar a vícios e a resultados que fogem da realidade. Os conceitos envolvidos na obtenção de amostras que produzam resultados confiáveis e livres de vícios não são triviais e precisam ser estabelecidos para o uso científico dos processos amostrais. Desse modo, torna-se necessário um estudo prévio da teoria da amostragem descrevendo as suas principais propriedades, as vantagens e desvantagens dos vários métodos existentes utilizados para a obtenção das amostras.

O presente trabalho, excluindo a introdução e conclusão, encontra-se dividido em 3 capítulos, assim distribuídos: no capítulo 2 será apresentada uma revisão bibliográfica, os objetivos deste trabalho e fatos históricos da estatística com seus principais contribuidores internacionais e nacionais, juntamente com os aspectos históricos que marcaram a amostragem.

No capítulo 3, a primeira seção é destinada às definições de vários termos técnicos que serão usados no corpo deste trabalho. Na seção 2 estuda-se as técnicas de amostragem pertinentes à Amostra Aleatória Simples (*AAS*) seguida da amostragem estratificada (*AE*), objetivo maior deste estudo. Nas seções seguintes descrevem-se breves comentários sobre a Amostragem por Conglomerados em estágio único e vários estágios, além da amostragem sistemática. Finalizando este capítulo, é apresentado um exemplo de aplicação que faz comparações entre as técnicas de amostragem utilizadas *AAS* e *AE*.

No capítulo 4, denominado "Aplicações com Dados Reais", é descrito em pormenores o projeto de extensão "Dimensionamento da população canina de Rio Claro-SP" e em seguida apresentado os resultados gerados pelas entrevistas domiciliares para este projeto.

Ao final, nos apêndices, tem-se os dados brutos que foram utilizados no exemplo de aplicação e os dados condensados da população canina, frutos do projeto de extensão,

separados por sexo.

2 Fundamentos

2.1 Revisão Bibliográfica

Os fatos históricos descritos na Seção 2.3, contemplando diversos aspectos desde a sua origem até os dias atuais, seja na estatística ou na amostragem, provêm principalmente de Memória (2004) complementado com informações provenientes do portal da Estatística da UFRN.

Na sequência, estudam-se as técnicas de amostragem, apresentadas no capítulo 3, tendo como principais fontes de referência Bolfarine e Bussab (2005) e Cochran (1965), com informações complementares e exemplos de aplicação provenientes de Muniz e Abreu (1999). Foram também consultados em várias instâncias Magalhães e Lima (2011) e Bussab e Moretin (2013), de onde provêm várias das definições utilizadas neste capítulo.

No tocante ao embasamento teórico, procedeu-se a leitura de vários artigos sobre o levantamento da população canina, onde se torna evidente a importância do conhecimento da população desses animais e o controle de suas zoonoses e suas relações com o ser humano, conforme podem ser verificados em Matos et al. (2002); Alves et al. (2005) e Soto et al. (2006). Foram consultados também alguns artigos relativos às técnicas de estratificação destacando-se entre eles Dalenius e Hodges (1959); Gunning, Horgan e Keogh (2004); Brito et al. (2010) e Horgan (2010), que tratam de alguns métodos com o intuito de estabelecer a escolha adequada dos limites de estratificação.

Em Dalenius e Hodges (1959) os autores propõem uma solução para o problema da determinação dos estratos em populações finitas considerando que a população é representada por uma função densidade de probabilidade $f_X(x)$ considerada, com média μ e variância σ^2 .

Em Gunning, Horgan e Keogh (2004), uma outra proposta é apresentada, onde a população é dividida em H estratos, considerando quaisquer limites para o intervalo de variação de X . Nesse método são feitas duas suposições, uma com base em Dalenius e Hodges (1959) e outra com base em Lavalée - Hidiroglou (1988 apud GUNNING; HORGAN; KEOGH, 2004).

Brito et al. (2010) propõe um algoritmo exato para resolver problemas na estratificação ótima com amostragem aleatória simples sem reposição usando alocação

proporcional. Esse método é sugerido com base em caminhos mínimos em grafos.

No artigo de Horgan (2010) são comparados diversos métodos utilizados para a obtenção dos limites ótimos da estratificação, tais como os métodos anteriormente já citados entre outros e sugere algumas modificações visando melhorias para o método da Estratificação Geométrica. O autor comenta ainda, que em populações assimétricas, a estratificação pode levar a uma melhoria substancial na precisão dos estimadores da amostra.

2.2 Objetivos

Estudo das técnicas de amostragem utilizadas em pesquisas de levantamento de dados que visam a estimação pontual e por intervalos de médias, proporções e totais; estudo das propriedades dos estimadores, vantagens e desvantagens de cada técnica considerando as dificuldades de implementação, seu custo e a precisão dos estimadores; e finalmente definir o plano amostral adequado para a aplicação no dimensionamento e classificação da população canina no município de Rio Claro - SP .

Em grande parte do mundo e também nos municípios brasileiros, a população de cães representa um sério problema de saúde pública, seja pela transmissão de doenças, seja pelo risco de agressão por mordidas (principalmente em crianças) ou até mesmo pelos casos de acidentes de trânsito provocados pela presença desses animais soltos nas vias públicas das cidades e nas estradas, conforme mencionado em WHO/WSPA (1992 apud MATOS et al. 2002).

Conhecer a estrutura e o tamanho da população canina, em qualquer localidade, é fator importantíssimo para o adequado planejamento e tomada de decisões para os programas de controle das zoonoses causadas por esses animais, destacando-se entre elas a raiva e a leishmaniose visceral.

Propor o método de amostragem que melhor se adapte, no que se refere a apresentação dos melhores resultados, para a aplicação no dimensionamento e classificação da referida população, seguindo a metodologia proposta em Matos et al. (2002).

2.3 Panorama Histórico

Nesta seção serão apresentados alguns dados relevantes à história da Estatística bem como sua evolução ao longo do tempo juntamente com seus principais pensadores. Seguidamente far-se-á um relato sobre a história da amostragem e seus contribuidores.

2.3.1 Da Estatística

A Estatística é uma ciência que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados e para utilização dos mesmos na tomada de

decisões. A palavra Estatística é derivada do termo latino "status" que significa estado.

Assim como a Matemática teve suas origens no convívio social, na necessidade da contagem, assim também nasce a Estatística. Há indícios de que há mais de 2000 anos A. C. já se faziam censos na Babilônia, China e Egito, fato este registrado até mesmo na Bíblia, no Antigo Testamento,¹ onde Moisés é instruído a fazer um levantamento de homens de Israel aptos para guerrear.

Assim, os povos da antiguidade faziam uso de técnicas para registrarem o número de seus habitantes, dos nascimentos, dos óbitos, faziam estimativas de riquezas individual e social, distribuição de terras entre os povos, cobranças de impostos entre várias outras coisas. Mas, foi a partir do século XVIII, que começaram a surgir as primeiras análises sistemáticas de fatos sociais, originando, assim, as primeiras tábuas, tabelas e números relativos.

Com as mutações ocorridas em toda a humanidade no decorrer de todos estes anos, o estudo de tais fatos foi adquirindo, aos poucos, feição verdadeiramente científica, surgindo, assim, a ciência responsável por esta área de estudo, hoje conhecida como **Estatística**. É devido a **Godofredo Achenwall**, (1719-1772), mais precisamente em 1748, na Alemanha, que essa ciência recebeu o seu verdadeiro nome, onde Achenwall determinou o seu objetivo e as relações com as outras ciências. Naquela época a Estatística era considerada a aritmética do Estado "Political Arithmetic".

Com o surgimento do Cálculo das Probabilidades, as tabelas tornaram-se mais complexas, surgiram as representações gráficas, e a Estatística deixou de ser simples catalogação de dados numéricos para se tornar o estudo de **como chegar as conclusões do todo - População, partindo das observações de partes desse todo - Amostras**. Data dos fins do século XIX, com trabalhos de Francis Galton (1822-1911), Karl Pearson (1857 - 1936) e William S. Gosset (1876-1937), conhecido com o pseudônimo de **Student**, o desenvolvimento da Estatística Matemática e suas aplicações.

Porém, foi no período de 1925 a 1960, considerada a **época áurea do pensamento estatístico**, que essa ciência foi enriquecida com a maior parte dos trabalhos sobre inferência de Ronald A. Fisher (1890 - 1962), Jerzy Neyman (1884 - 1981), Egon S. Pearson (1895 - 1980) e Abraham Wald (1902 - 1950) entre outros, além do desenvolvimento dos delineamentos experimentais e levantamentos por amostragem, assim como as ideias fundamentais sobre séries temporais, análise multidimensional e as contribuições bayesianas. Apesar dos importantes trabalhos, publicados entre 1960 e 1981, esse período foi basicamente de consolidação das ideias anteriormente desenvolvidas. Dentre esses grandes pensadores, destaca-se **J. Neyman**, considerado um dos gigantes da Estatística.

Essa ciência veio a ganhar um grande impulso nas décadas de 1940 e 1950 com o surgimento das máquinas de calcular, manuais e elétricas, tornando-se ferramentas comuns e chegando a era atual como uma ferramenta muito poderosa com seus processos e

¹Ver Bíblia Sagrada: Antigo Testamento; Números 1, 1-3.

técnicas, graças a sua matematização e a influência crescente do uso dos computadores juntamente com seus aplicativos, contribuindo, assim, para a organização dos negócios e recursos do mundo moderno.

No Brasil, podemos dizer que a Estatística tem sua história associada à história do Instituto Brasileiro de Geografia e Estatística - IBGE, criado em 1934, passando a existir de fato apenas em 1936, quando os serviços geográficos foram a ele vinculados. Hoje, esse instituto é denominado de Fundação Instituto Brasileiro de Geografia e Estatística - IBGE, integrante da administração federal, cujas finalidades básicas são: "a pesquisa, a produção, a análise e a difusão de informações e estudos de natureza estatística, geografia, cartografia, geodésica, demografia e sócio-econômica, de recursos naturais e de condições do meio ambiente, necessárias ao conhecimento da realidade física, humana, econômica e social, com vistas, especialmente, à execução de programas e projetos de desenvolvimento nacional".

Em Julho de 1955, Fisher, em companhia de C. Radhakrishna Rao, veio ao Brasil participar do 2º Congresso Internacional de Biometria, realizado em Campinas, e, em 1961, Neyman permaneceu por um mês em São Paulo, onde por ocasião de sua visita sugeriu a criação de um Departamento de Estatística junto a Universidade de São Paulo, fato este, que só foi concretizado onze anos depois, em 1972, quando se deu de fato a criação desse departamento juntamente com o curso de Bacharelado em Estatística.

2.3.2 Da Amostragem

Há indícios de que, os levantamentos por amostragem, tiveram suas origens desde os primórdios da Estatística. Porém, o seu desenvolvimento só aconteceu de fato após as contribuições da teoria da Estatística, proveniente principalmente dos trabalhos de Neyman e Pearson, no século XX, em particular, na década de 1930.

Entre os países que mais se destacaram nos levantamentos por amostragem, pode-se citar os Estados Unidos, nas atividades exercidas pelo "Bureau of the Census" e a Índia, que utilizou o método com a finalidade do planejamento econômico de 1930, sob orientação do Estatuto Indiano de Estatística. Outra aplicação importante desse método, utilizado em diversos países, é na Agricultura, principalmente nos diversos levantamentos e nos Censos Agrícolas.

O levantamento por amostragem é praticamente o único método utilizado nas pesquisas econômicas e sociais, sendo que, na prática, raramente uma amostra é selecionada de modo estritamente aleatório.

Conforme pode-se notar em Memoria (2004), deve-se principalmente, a Arthur Lyon Bowley (1869 - 1957), a teoria e a possibilidade de aplicações práticas do método representativo. Bowley considerou a amostra representativa como podendo se realizar por dois métodos: seleção intencional "purposive" ou ao acaso "amostra aleatória", podendo haver combinação dos dois. No caso da amostra intencional, ela se forma

mediante uma seleção de grupos que se supõe dão à amostra as mesmas características que tem a população. Esse método era aconselhado quando não era possível a seleção ao acaso, sendo, às vezes, mais econômico, porém com a desvantagem de não poder ser calculado o erro de amostragem. Outra aplicação da amostragem representativa foi feita pelos estatísticos italianos Corrado Gine (1884 - 1965) e seu assistente Luigi Galvani, para obterem os dados do Censo de 1921.

Neyman (1934 apud MEMÓRIA, 2004) mostrou em seu artigo a superioridade da amostragem aleatória estratificada sobre a seleção intencional, então aconselhou como método representativo, apesar de ter sido severamente criticado por Gine e Galvani. A descrença no método representativo, argumentou Neyman, devia ser na amostragem intencional e não na amostragem probabilística, que ele recomendou como a do método representativo. Na verdade, não há processo seguro para se saber se determinada amostra é representativa de uma população. Operacionalmente, define-se como representativa uma amostra selecionada ao acaso, de modo que cada unidade de amostragem tenha uma probabilidade conhecida, e diferente de zero, de participar da amostra.

Outra importante contribuição de Neyman foi seu artigo Neyman (1938 apud MEMÓRIA, 2004) sobre amostragem dupla, em duas fases. Esse artigo, juntamente com o artigo de Neyman (1952 apud MEMÓRIA, 2004), seriam suficientes para garantir a Neyman o reconhecimento de sua preeminência na teoria dos levantamentos por amostragem. Destaca-se ainda, a análise estatística dos resultados de ensaios para provocar "chuvas artificiais", que deu origem ao artigo de Neyman (1967 apud MEMÓRIA, 2004).

Merece destaque também, a participação de Fisher, que apesar de não ter escrito qualquer livro sobre as técnicas de amostragem, especialmente nas aplicações em levantamentos, muitas ideias nessa teoria, são oriundas de seus trabalhos, como pode-se citar a casualização e controle local (estratificação). Fisher estudou também o uso de amostragem em parcelas experimentais, com consequências no desenvolvimento e melhoramento das estimativas da produção agrícola e de áreas cultivadas.

Segundo consta, Leslie Kish (1910 - 1999) publicou vários trabalhos, entre eles um dos primeiros com seu chefe Roe Goodman, (GOODMAN; KISH, 1950 apud MEMÓRIA, 2004). Ele executou levantamentos por amostragem em vários países, principalmente na América do Sul. Por sua fluência em espanhol, o livro de Kish, *Survey Sampling*, foi editado em 1965, em espanhol, e traduzido para várias línguas incluindo o chinês.

No Brasil, a amostragem começou a ser usada a partir do Censo Demográfico de 1950, na avaliação das tabulações avançadas, que se fizeram necessárias dada a demora da publicação dos resultados obtidos no censo completo.

Um exemplo dessas atividades aplicadas no Brasil é a Pesquisa Nacional de Amostra de Domicílios - PNAD - implantada progressivamente a partir de 1967, para a obtenção de informações básicas necessárias para o estudo do desenvolvimento econômico do

país, que se estendeu até 1970, com seus resultados apresentados com periodicidade trimestral. A partir daí, essa pesquisa foi interrompida, sendo substituída pelos censos demográficos decenais que se sucederam após as décadas de 1970. Na época, os principais temas investigados pela pesquisa, além das características gerais da população, da educação, do rendimento e da habitação, foram a migração e a fecundidade.

Nesse contexto, destaca-se também as contribuições de William Madow (1895 - 1973), que sendo professor da Universidade de São Paulo, em 1946 e 1947, ministrou durante o período de férias acadêmicas, de dezembro de 1946 a fevereiro de 1947, um curso intensivo sobre a teoria dos levantamentos por amostragem, no Rio de Janeiro, a técnicos brasileiros de Estatística, a convite do IBGE.

Atualmente, no cenário nacional, destaca-se as relevantes contribuições dos professores e pesquisadores Heleno Bolfarine e Wilton de Oliveira Bussab (1940 - 2011). Heleno Bolfarine e Wilton de Oliveira Bussab, juntos, são autores do livro "Elementos de Amostragem" (BOLFARINE; BUSSAB, 2005), os quais chegaram a ser ganhadores do prêmio Jabuti 2006 com essa obra. A mesma representa a literatura base para a elaboração do presente trabalho.

O professor Heleno Bolfarine, com título de Ph.D em Estatística pela Universidade da Califórnia, Berkeley, é atualmente professor titular do Instituto de Matemática e Estatística da USP. É pesquisador, atuando principalmente, nas linhas de pesquisa de Modelo de Regressão e Aplicações na USP (líder), Inferência Estatística na UFSCar, Análise Bayesiana, entre outras. Foi vencedor do prêmio ABE 2012 outorgada pela Associação Brasileira de Estatística.

O professor Wilton de Oliveira Bussab, foi um dos primeiros mestres da USP (1971) e fez seu doutorado na London School of Economics (Universidade de Londres) em 1976. Foi professor do Instituto de Matemática e Estatística da USP, UFSCar (1977 - 1981) e da escola de Administração de Empresas da Fundação Getúlio Vargas (FGV), em São Paulo, exercendo, ainda, o cargo de Diretor Acadêmico desta fundação por mais de 40 anos. Foi Presidente da Associação Brasileira de Estatística (ABE) em duas ocasiões, de 1986 a 1988 e de 2006 a 2008. Desenvolveu também inúmeras atividades de assessoria estatística para empresas e instituições governamentais, dentre elas o IBGE. Tinha grande experiência nas áreas de análise multivariada, amostragem e metodologia quantitativa. Dos seus vários livros publicados destaca-se "Estatística Básica" em co-autoria com Pedro A. Moretim (BUSSAB; MORETIM, 2013).

3 Técnicas de Amostragem

No presente capítulo serão descritas as duas principais técnicas de amostragem, de natureza probabilística, apresentando seus principais estimadores e comparando alguns desses métodos para escolha, do mais adequado, para o dimensionamento da população de cães do município de Rio Claro, estado de São Paulo.

Na amostragem probabilística, os elementos componentes de uma amostra são escolhidos aleatoriamente. Tal procedimento tem por finalidade evitar qualquer tendenciosidade, que jamais deverá existir nas investigações estatísticas por amostragem. Geralmente, neste tipo de amostragem é exigido um cadastro ou uma listagem dos elementos da população (Universo de Estudo).

Dentro da amostragem probabilística, destacam-se: a Amostragem Aleatória Simples (AAS); a Amostragem Estratificada (AE); a Amostragem por Conglomerados (AC); a Amostragem em Múltiplos Estágios (AME) e a Amostragem Sistemática (AS), as quais serão detalhadas nas seções seguintes.

Existem ainda outros dois tipos de amostragem que são a Amostragem por Quotas e a Amostragem Intencional. A amostragem por quotas é muito usada na realização de pesquisas tais como as pesquisas de mercado e as de opinião pública para fins de predição eleitoral. O método, apesar de não ter caráter científico, haja vista não ser de natureza probabilística, é essencialmente prático. O dimensionamento de uma amostra é feito apenas com base numa quota percentual, levando em conta não só a magnitude do "universo" estudado, como também a natureza do fenômeno observado numericamente.

Já o método de amostragem intencional é pouco utilizado na realização de pesquisas estatísticas, tendo em vista o mesmo não ter caráter científico. Nesse método, uma amostra não é dimensionada matematicamente. O número de elementos que formarão a amostra é fixado de modo arbitrário pelo pesquisador e os seus elementos, conforme o próprio nome sugere, são selecionados "intencionalmente" segundo o gosto e a preferência do mesmo. Para a aplicação do referido método não é exigido um cadastro ou listagem do universo a ser estudado.

3.1 Definições

Antes de iniciar o estudo das técnicas de amostragem de natureza probabilística, serão definidos alguns termos, os quais serão utilizados no desenvolvimento deste trabalho.

Definição 3.1. *Define-se População ou Universo de Estudo como sendo o conjunto de entes portadores de, pelo menos, uma característica comum.*

Notação: $U = \{1, 2, \dots, N\}$.

Definição 3.2. *Define-se Amostra ordenada como uma sequência finita qualquer de n unidades de U , ou seja,*

$$s = (k_1, \dots, k_n)$$

tal que

$$k_i \in U \quad e \quad n \leq N.$$

O termo k_i é chamado de i -ésimo componente de s .

Definição 3.3. **Amostragem** *é o processo de seleção de uma amostra, que possibilita o estudo das características de uma população.*

Definição 3.4. *Define-se como fração amostral o valor obtido pela fração $\frac{n}{N}$ onde n representa o número de elementos de uma amostra e N o número de elementos da população.*

Definição 3.5. *Segundo Muniz e Abreu (1999), quando a fração de amostragem $\frac{n}{N} < 0,05$, ou seja, a amostra tem menos de 5% da população, a população é considerada infinita. Se $\frac{n}{N} \geq 0,05$ a população é considerada finita.*

Definição 3.6. *Um conjunto X se diz enumerável, se X é finito ou existe uma bijeção $\varphi : \mathbb{N} \rightarrow X$, onde \mathbb{N} representa o conjunto dos números naturais.*

Definição 3.7. *Define-se como espaço amostral o conjunto de todos os resultados possíveis de um fenômeno aleatório.*

Definição 3.8. *Define-se uma quantidade Y , associada a cada resultado possível do espaço amostral como variável aleatória discreta, assumindo valores pertencentes a um conjunto enumerável. A variável é contínua quando assume valores em um intervalo de números reais.*

Definição 3.9. *Define-se como Função de Probabilidade a função que associa a cada valor assumido pela variável aleatória discreta a probabilidade do evento correspondente, satisfazendo $P(X = x) \geq 0$ e $\sum_x P(X = x) = 1$.*

Notação: $P(X = x_i)$

Definição 3.10. Diz-se que $f(x)$ é uma **função densidade de probabilidade** para uma variável aleatória contínua X , se satisfaz as condições:

i. $f(x) \geq 0$, para todo $x \in (-\infty, \infty)$;

ii. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Definição 3.11. Define-se como **Esperança ou Esperança Matemática** uma média aritmética ponderada. A Esperança Matemática é um número real.

Para uma variável aleatória discreta Y , com função de probabilidade $P(Y = y_i)$, tem-se

$$E(Y) = \sum_{i=1}^N Y_i P(Y = y_i) \quad e \quad E[h(y)] = \sum_{i=1}^N h(y_i)P(Y = y_i),$$

e para uma variável aleatória contínua Y , com função densidade de probabilidade $f(y)$, tem-se

$$E(Y) = \int_{-\infty}^{\infty} y \cdot f(y)dy \quad e \quad E[h(y)] = \int_{-\infty}^{\infty} h(y)f(y)dy.$$

Definição 3.12. Define-se como **Variância** a medida que dá o grau de dispersão (ou concentração) de Y em torno da média.

$$V(Y) = E(Y)^2 - [E(Y)]^2.$$

Definição 3.13. Define-se como **Parâmetro** um valor desconhecido associado a uma característica da população. São exemplos de parâmetros: a média (μ), a variância (σ^2), a proporção (P), entre outras.

Definição 3.14. Define-se **estimador** como a função que fornece um valor de um parâmetro baseando-se nas observações de uma amostra.

Definição 3.15. Um estimador $\hat{\theta}$ é **não viciado** para o parâmetro θ , se

$$E(\hat{\theta}) = \theta.$$

Definição 3.16. Um estimador $\hat{\theta}$ é **consistente** para o parâmetro θ , se

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta \quad e \quad \lim_{n \rightarrow \infty} V(\hat{\theta}) = 0.$$

Definição 3.17. Considerando dois estimadores não viciados $\hat{\theta}_1$ e $\hat{\theta}_2$, dizemos que $\hat{\theta}_1$ é **mais eficiente** que $\hat{\theta}_2$, se

$$V(\hat{\theta}_1) < V(\hat{\theta}_2).$$

Definição 3.18. Seja $f_i(s)$ a variável que indica a frequência que a i -ésima unidade populacional aparece na amostra s . Indica-se por $\delta_i(s)$ a variável que informa a presença da i -ésima unidade na amostra s , isto é,

$$\delta_i(s) = \begin{cases} 1, & \text{se } i \in s \\ 0, & \text{se } i \notin s. \end{cases}$$

Definição 3.19. *Define-se como **Plano amostral** o conjunto de etapas que descreve todos os métodos e medidas envolvidos na execução da amostragem. O plano amostral contém todas as informações, como o método usado para seleção da amostra, erro de estimação, tamanho da amostra, estimadores usados na pesquisa, entre outros.*

Definição 3.20. *Define-se como **Planos Simétricos** os planos nos quais as esperanças e variâncias são as mesmas para todas as variáveis.*

3.2 Amostra Aleatória Simples - AAS

A amostragem aleatória simples é um processo para selecionar uma amostra de tamanho n retirada do universo populacional U , que contém um número finito N de unidades, de modo que cada um dos n elementos tenham igual probabilidade de ser selecionados. Esse método é o mais simples e o mais importante para a seleção de uma amostra. O seu emprego é recomendado para o estudo de populações homogêneas, isto é, que o valor das medidas variem pouco de uma unidade para outra.

O método apresenta a vantagem de ser bastante prático, motivo pelo qual é o mais usado, universalmente, dentre os demais métodos existentes, de natureza probabilística. Como desvantagem, pode-se citar o custo operacional elevado das pesquisas estatísticas, principalmente, quando o universo estudado apresentar grande variabilidade ou então achar-se contido numa vasta área territorial.

Esse plano caracteriza-se pela existência de um sistema de referências completo, descrevendo cada uma das unidades elementares, ou seja,

$$U = \{1, 2, \dots, N\}.$$

Para a utilização desse plano, e, dispondo-se de um dispositivo aleatório qualquer, sorteia-se com igual probabilidade um elemento x da população U . Sendo permitido o sorteio de uma mesma unidade mais de uma vez, tem-se o processo AAS com reposição, designado por AAS_C . Se o elemento sorteado é removido do universo populacional antes do sorteio do próximo número, tem-se a AAS sem reposição, indicado por AAS_S .

As duas próximas sub-seções tratarão cada um desses planos separadamente.

3.2.1 Amostragem Aleatória Simples com Reposição - AAS_C

Esse plano opera da seguinte forma:

- A população é numerada de 1 a N , ou seja,

$$U = \{1, 2, \dots, N\};$$

- Sorteia-se, com igual probabilidade, uma das N unidades de U ;

- Repõe-se esta unidade na população U e sorteia-se o elemento seguinte;
- Repete-se o processo anterior até que sejam sorteadas as n unidades, número este prefixado inicialmente, os quais corresponderão aos elementos pertencentes a amostra.

Nesse plano amostral, a variável f_i , conforme Definição 3.18, página 29, segue uma distribuição binomial com parâmetros n e $\frac{1}{N}$, denotada por

$$f_i \sim b\left(n, \frac{1}{N}\right),$$

de modo que

$$E(f_i) = \frac{n}{N} \quad e \quad V(f_i) = \frac{n}{N} \left(1 - \frac{n}{N}\right)$$

Definição 3.21. *Os experimentos nos quais a variável aleatória em questão assume somente dois valores, 1 ou 0, são chamados de **ensaios de Bernoulli**. É comum denominar essas respostas de **sucesso** e **fracasso** com probabilidade de ocorrer sucesso igual a p .*

$$Y = \begin{cases} 1, & \text{se ocorre sucesso com probabilidade } p \\ 0, & \text{se ocorre fracasso com probabilidade } 1 - p. \end{cases}$$

A função de probabilidade é dada por

$$P(Y = y) = p^y(1 - p)^{1-y}.$$

Definição 3.22. *Seja Y uma variável aleatória que representa o número de "sucessos" em n ensaios independentes de Bernoulli com probabilidade de sucesso p . A variável aleatória que conta o número total de "sucessos" é denominada **Binomial** com parâmetros n e p e sua função de probabilidade é dada por*

$$P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k},$$

para $k = 0, 1, \dots, n$.

3.2.1.1 Estimadores para o total e a média da população

Teorema 3.1. *Seja Y uma variável aleatória com $E(Y) = \mu$ e $V(Y) = \sigma^2$. A média amostral dada por*

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n Y_i,$$

a qual é um estimador não viciado e consistente para a média populacional μ sendo sua variância $\frac{\sigma^2}{n}$.

Corolário 3.1. Seja τ , o total populacional para uma variável aleatória Y , dado por $\tau = \sum_{i=1}^N Y_i$. Dentro do plano AAS_C ,

$$\hat{\tau} = N \bar{y} = \frac{N}{n} \sum_{i=1}^n Y_i$$

é um estimador não viciado para τ com

$$V(\hat{\tau}) = V(N\bar{y}) = N^2 V(\bar{y}) = N^2 \frac{\sigma^2}{n}.$$

$\hat{\tau}$ é denominado estimador expansão do total populacional. Note que

$$\tau = \sum_{i \in s} Y_i + \sum_{i \notin s} Y_i,$$

e $\hat{\tau}$ pode ser escrito como

$$\hat{\tau} = \frac{N}{n} \sum_{i \in s} Y_i = n\bar{y} + (N - n)\bar{y},$$

de modo que $(N - n)\bar{y}$ estima a parte não observada $\sum_{i \notin s} Y_i$ de τ .

3.2.1.2 Estimação da Variância Populacional

Teorema 3.2. Seja σ^2 a variância populacional para uma variável aleatória Y , com $E(Y) = \mu$, dada por

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

No plano AAS_C ,

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

é um estimador não viciado para $V(Y) = \sigma^2$

Demonstração. Note que

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

de modo que

$$E[(n-1) s^2] = E \left(\sum_{i=1}^n Y_i^2 \right) - \frac{1}{n} E \left(\sum_{i=1}^n Y_i \right)^2.$$

Para a classe de planos amostrais simétricos e de tamanho n fixo e com f_i , conforme Definição 3.18, página 29, tem-se que

$$\begin{aligned} E\left(\sum_{i \in s} Y_i^2\right) &= E\left(\sum_{i=1}^N f_i Y_i^2\right) = \sum_{i=1}^N Y_i^2 E(f_i) \\ &= E(f) \sum_{i=1}^N Y_i^2 = \frac{n}{N} \sum_{i=1}^N Y_i^2 \\ &= \frac{n}{N} \left[\sum_{i=1}^N Y_i^2 - N\mu^2 + N\mu^2 \right] \\ &= \frac{n}{N} \left[\sum_{i=1}^N Y_i^2 - 2N\mu^2 + N\mu^2 + N\mu^2 \right]. \end{aligned}$$

Substituindo $N\mu$ por $\sum_{i=1}^N Y_i$, tem-se:

$$\begin{aligned} E\left(\sum_{i \in s} Y_i^2\right) &= \frac{n}{N} \left[\sum_{i=1}^N Y_i^2 - 2\mu \sum_{i=1}^N Y_i + N\mu^2 + N\mu^2 \right] \\ &= \frac{n}{N} \left[\sum_{i=1}^N (Y_i - \mu)^2 + N\mu^2 \right]. \end{aligned}$$

Como $\sigma^2 = \frac{\sum_{i=1}^N (Y_i - \mu)^2}{N}$ então $N\sigma^2 = \sum_{i=1}^N (Y_i - \mu)^2$, e daí resulta:

$$\begin{aligned} E\left(\sum_{i \in s} Y_i^2\right) &= \frac{n}{N} [N\sigma^2 + N\mu^2] \\ &= n\sigma^2 + n\mu^2 \\ &= E(f_i) [N\sigma^2 + N\mu^2] \\ &= E(f_i) N (\sigma^2 + \mu^2) = \frac{n}{N} N (\sigma^2 + \mu^2) = n\sigma^2 + n\mu^2, \\ E[(n-1)s^2] &= n\sigma^2 + n\mu^2 - \frac{1}{n} E\left(\sum_{i=1}^n Y_i\right)^2. \end{aligned}$$

Mas $V(\sum Y_i) = E(\sum Y_i)^2 - [E(\sum Y_i)]^2$. Assim:

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

□

Com os resultados anteriores, tem-se estimadores não viciados para as variâncias dos estimadores de μ e τ , que estão condensados no corolário a seguir:

Corolário 3.2. *Para o plano AAS_C, tem-se que*

$$\widehat{V}(\bar{y}) = \frac{s^2}{n}$$

é um estimador não viciado para $V(\bar{y}) = \frac{\sigma^2}{n}$ e

$$\widehat{V}(\widehat{\tau}) = N^2 \frac{s^2}{n}$$

é um estimador não viciado para $V(\widehat{\tau}) = \frac{N^2 \sigma^2}{n}$.

3.2.1.3 Teorema Central do Limite - TCL

Conforme apresentado em Magalhães e Lima (2011), tem-se:

Teorema 3.3. *Suponha uma amostra aleatória simples de tamanho n retirada de uma população com média μ e variância σ^2 . Representando tal amostra por n variáveis aleatórias independentes (X_1, \dots, X_n) e, denotando sua média por \bar{X} , tem-se que*

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{n \rightarrow \infty} N(0, 1).$$

Este teorema garante que, quando o tamanho de uma amostra aumenta, ou seja para n grande, independente da forma da distribuição da população, a distribuição amostral de \bar{X} aproxima-se cada vez mais de uma distribuição normal.

A demonstração desse teorema será omitida neste trabalho, tendo-se em vista a exigência de técnicas e recursos bem mais avançados para este fim, conforme citado em Magalhães e Lima (2011, p. 238) e Bussab e Moretin (2013).

3.2.1.4 Normalidade Assintótica e Intervalos de Confiança - IC

Pelo TCL, para $n \rightarrow \infty$, tem-se para AAS_C, que:

$$Z = \frac{\bar{y} - \mu}{\frac{\sigma}{\sqrt{n}}} \cong N(0, 1)$$

e

$$Z_{\widehat{\tau}} = \frac{\widehat{\tau} - \tau}{N \frac{\sigma}{\sqrt{n}}} \cong N(0, 1).$$

Disso, tem-se os Intervalos de confiança para μ e τ , com σ^2 conhecida e com coeficientes de confiança $(1 - \alpha)\%$, onde:

$$IC(\mu) = \bar{y} \pm z_{(\alpha/2)} \frac{\sigma}{\sqrt{n}} \quad e \quad IC(\tau) = \widehat{\tau} \pm z_{(\alpha/2)} \frac{N\sigma}{\sqrt{n}},$$

sendo $z_{\alpha/2}$ o ponto que deixa $\frac{\alpha}{2}\%$ de área à direita.

O termo $z_{(\alpha/2)} \frac{\sigma}{\sqrt{n}}$, será indicado por ε e o termo $z_{(\alpha/2)} \frac{N\sigma}{\sqrt{n}}$, indicado por ε_τ representam os erros máximos de estimação.

3.2.1.5 Determinação do Tamanho da Amostra

De forma que o erro máximo de estimação seja ε , com probabilidade $1 - \alpha$, tem-se:

$$P(|\bar{y} - \mu| \leq \varepsilon) \cong 1 - \alpha$$

$$P\left(|\bar{y} - \mu| \leq z_{(\alpha/2)} \frac{\sigma}{\sqrt{n}}\right) \cong 1 - \alpha$$

Para ε fixado, a solução para o problema consiste em se determinar n tal que:

$$\varepsilon = z_{(\alpha/2)} \frac{\sigma}{\sqrt{n}}$$

ou

$$\frac{\varepsilon^2}{z_{(\alpha/2)}^2} = \frac{\sigma^2}{n} \implies n = \frac{\sigma^2}{\left(\frac{\varepsilon}{z_{(\alpha/2)}}\right)^2} = \frac{\sigma^2 z_{(\alpha/2)}^2}{\varepsilon^2}.$$

Para a determinação do tamanho da amostra, é preciso fixar o erro máximo desejado ε com um determinado grau de confiança, atribuído pelo pesquisador, além de ter um conhecimento prévio sobre a estimativa da variabilidade da população ($\hat{\sigma}^2$), que pode ser obtida através do conhecimento de pesquisas anteriores e de uma amostra piloto. No caso de distribuições simétricas ($\cong Normal$) pode-se fazer a amplitude do intervalo de confiança igual a 4σ , que contém 95% dos elementos da população.

3.2.1.6 Estimação de Proporções

Nos levantamentos por amostragem, quando deseja-se estimar a proporção de unidades da população que apresentam uma determinada característica, associa-se a cada elemento dessa população os valores 1 e 0, conforme esta unidade apresenta a característica desejada ou não. Assim tem-se a variável aleatória Y , com $E(Y) = \mu$ e $V(Y) = \sigma^2$, definida por:

$$Y_i = \begin{cases} 1, & \text{se o elemento } i \text{ possui a característica,} \\ 0, & \text{caso contrário.} \end{cases}$$

Então,

$$P = \frac{1}{N} \sum_{i=1}^N Y_i = \mu$$

é a proporção de unidades na população que possuem a característica de interesse.

Desde que Y_i assumam apenas os valores 1 e 0, pode-se escrever:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - P)^2 = P(1 - P).$$

Seja s uma amostra observada de tamanho n e seja m o número de elementos da amostra que possuem a característica específica. Do Teorema 3.1, página 31, tem-se para a AAS_C , que um estimador não viciado para P é dado por

$$\hat{P} = \bar{y} = \frac{1}{N} \sum_{i \in s} Y_i = \frac{m}{n}$$

e que

$$V(\hat{P}) = \frac{\sigma^2}{n} = \frac{PQ}{n},$$

sendo $Q = 1 - P$.

Do Teorema 3.2, página 32, tem-se que um estimador não viciado de σ^2 é dado por

$$s^2 = \frac{n}{n-1} \hat{P}\hat{Q},$$

sendo $\hat{Q} = 1 - \hat{P}$.

Consequentemente tem-se que um estimador não viciado para $V(\hat{P})$ é dado por

$$\hat{V}(\hat{P}) = \frac{s^2}{n} = \frac{\hat{P}\hat{Q}}{n-1}.$$

Um intervalo de confiança aproximado para P é dado por

$$\left(\hat{P} - z_{(\alpha/2)} \sqrt{\frac{\hat{P}\hat{Q}}{n-1}}; \hat{P} + z_{(\alpha/2)} \sqrt{\frac{\hat{P}\hat{Q}}{n-1}} \right).$$

Sendo PQ sempre menor ou igual a $\frac{1}{4}$, tem-se que um intervalo de confiança conservativo para P é dado por

$$\left(\hat{P} - z_{(\alpha/2)} \sqrt{\frac{1}{4(n-1)}}; \hat{P} + z_{(\alpha/2)} \sqrt{\frac{1}{4(n-1)}} \right).$$

3.2.2 Amostra Aleatória Simples sem Reposição - AAS_S

A Amostra Aleatória Simples sem Reposição - AAS_S opera da seguinte forma:

- A população é numerada de 1 a N , ou seja,

$$U = \{1, 2, \dots, N\};$$

- Sorteia-se, com igual probabilidade, uma das N unidades de U ;

- Sorteia-se o elemento seguinte, tendo o elemento anterior sido retirado da população;
- Repete-se o processo anterior até que sejam sorteadas as n unidades, número este prefixado inicialmente, os quais corresponderão aos elementos pertencentes a amostra.

Dessa forma, cada elemento da população pode aparecer uma única vez na amostra.

Teorema 3.4. *No plano Amostral AAS_S, a variável f_i , (Definição 3.18, página 29), segue à distribuição de **Bernoulli**, (ver Definição 3.21 página 31), com parâmetros $p = \frac{n}{N}$, denotado por*

$$f_i \sim Ber\left(\frac{n}{N}\right),$$

de forma que

$$E(f_i) = P \quad e \quad V(f_i) = P(1 - P)$$

satisfazendo:

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

Demonstração. Para a condição $P(f_i = 1) = \frac{n}{N}$ tem-se:

$$\begin{aligned} P(f_i = 1) &= \frac{1}{N} + \frac{N-1}{N} \frac{1}{N-1} + \frac{N-1}{N} \frac{N-2}{N-1} \frac{1}{N-2} + \dots \\ &+ \frac{N-1}{N} \frac{N-2}{N-1} \frac{N-3}{N-2} \dots \frac{N-n+1}{N-n+2} \frac{1}{N-n+1} \\ &= \underbrace{\frac{1}{N} + \frac{1}{N} + \frac{1}{N} + \dots + \frac{1}{N}}_n \\ &= \frac{n}{N}, \end{aligned}$$

com $i = 1, \dots, N$ □

A demonstração da condição $P(f_i = 0)$ segue de forma análoga.

3.2.2.1 Correção para Populações Finitas - CPF

Em uma amostra aleatória simples de tamanho n , proveniente de uma população infinita, tem-se que a variância de \bar{y} é $\frac{\sigma^2}{n}$. Quando a população for finita, esse valor modifica-se mediante a introdução do fator $\left(\frac{N-n}{N}\right)$, o qual é denominado de "correção das populações finitas - (CPF)".

Desde que a fração de amostragem, $\frac{n}{N}$, permaneça baixa, esses valores se aproximam da unidade. Na prática, as CPF's podem ser ignoradas quando a fração de amostragem não exceder a 5%, conforme encontra-se em Cochran (1965).

3.2.2.2 Estimação do Total e da Média Populacionais

Teorema 3.5. *Seja Y uma variável aleatória com $E(Y) = \mu$ e $V(Y) = \sigma^2$, então com relação à AAS_S , um estimador não viciado para o total populacional é*

$$\hat{\tau} = \frac{N}{n} \sum_{i=1}^n Y_i$$

apresentando as seguintes propriedades

1. $E(\hat{\tau}) = E\left(\frac{n}{N}\right) \sum_{i=1}^N Y_i = \frac{n}{N} \tau = n \mu$;
2. $V(\hat{\tau}) = \frac{n}{N} \left(1 - \frac{n}{N}\right) NS^2 = n\left(1 - \frac{n}{N}\right)S^2$,

sendo

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \mu)^2$$

a variância populacional.

Na AAS_S usa-se S^2 como variância populacional para garantir que $E(s^2) = S^2$. Se $\frac{n}{N} < 0,05$ e a população é considerada infinita, sendo essa substituição desnecessária.

Teorema 3.6. *Com relação à AAS_S , a média amostral*

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

é um estimador não viciado com variância dada por

$$V(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{S^2}{n}.$$

3.2.2.3 Estimação da Variância Populacional

Teorema 3.7. *A variância da amostra*

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

é um estimador não viciado no plano amostral AAS_S para

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \mu)^2.$$

Demonstração. A demonstração desse teorema é análoga ao Teorema 3.2, página 32.

□

Corolário 3.3. Para o plano amostral AAS_S , a estatística

$$\widehat{V}(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{s^2}{n}$$

é um estimador não viciado para $V(\bar{y})$ e

$$\widehat{V}(\widehat{\tau}) = N^2 \left(1 - \frac{n}{N}\right) \frac{s^2}{n}$$

é não viciado para $V(\widehat{\tau})$.

3.2.2.4 Normalidade Assintótica e Intervalos de Confiança

Todos os casos apresentados para a AAS_C , têm seu equivalente para a AAS_S , mudando apenas a expressão da variância amostral. Assim, para a AAS_S , tem-se:

$$\frac{\bar{y} - \mu}{\sqrt{(1 - n/N) s^2/n}} \cong N(0, 1) \implies IC(\mu) = \bar{y} \pm z_{(\alpha/2)} \sqrt{\frac{(1 - n/N) s^2}{n}}$$

e

$$\frac{\widehat{\tau} - \tau}{\sqrt{N^2(1 - n/N) (s^2/n)}} \cong N(0, 1) \implies IC(\tau) = \widehat{\tau} \pm z_{(\alpha/2)} \sqrt{\frac{N^2 (1 - n/N) s^2}{n}}.$$

3.2.2.5 Determinação do tamanho da amostra

De posse dos resultados apresentados para a AAS_C e fazendo-se as devidas adaptações para o caso da AAS_S , pode-se observar que:

$$V_{AAS_S}[\bar{y}] = \left(1 - \frac{n}{N}\right) \frac{S^2}{n} = \frac{S^2}{n / \left(1 - \frac{n}{N}\right)} = \frac{S^2}{n'},$$

sendo $n' = \frac{n}{1 - \frac{n}{N}}$.

Partindo-se da relação com ε e α fixados anteriormente, tem-se

$$P(|\bar{y} - \mu| \leq \varepsilon) \cong 1 - \alpha \implies P(-\varepsilon \leq \bar{y} - \mu \leq \varepsilon) \cong 1 - \alpha$$

$$P\left(\frac{-\varepsilon}{\sqrt{V(\bar{y})}} \leq \frac{\bar{y} - \mu}{\sqrt{V(\bar{y})}} \leq \frac{\varepsilon}{\sqrt{V(\bar{y})}}\right) \cong 1 - \alpha$$

$$P\left(\frac{-\varepsilon}{\sqrt{S^2/n'}} \leq Z \leq \frac{\varepsilon}{\sqrt{S^2/n'}}\right) \cong 1 - \alpha$$

$$\text{Se } z = \frac{\varepsilon}{\sqrt{S^2/n'}}, \text{ então } z^2 = \frac{\varepsilon^2}{S^2/n'} \text{ e } n' = \frac{z^2 S^2}{\varepsilon^2}.$$

Comparando as expressões de n' e isolando-se o valor de n , tem-se

$$\frac{n}{1 - \frac{n}{N}} = \frac{z^2 S^2}{\varepsilon^2} \implies n = \frac{S^2/K}{1 + \frac{S^2}{NK}},$$

sendo $K = \frac{\varepsilon^2}{z_{\alpha/2}^2}$, onde ε^2 representa o erro de estimação e $z = z_{\alpha/2}$ representa o percentil $\frac{\alpha}{2}$ da distribuição normal padrão.

Analogamente, o tamanho da amostra para que

$$P(|\hat{\tau} - \tau| \leq \varepsilon_\tau) = 1 - \alpha$$

é dado por

$$n = \frac{N}{\frac{K}{N S^2} + 1}.$$

3.2.2.6 Estimação de Proporções

De forma semelhante à AAS_C e associando-se a cada elemento dessa população os valores 1 e 0, conforme esta unidade apresenta a característica desejada ou não, e, desde que Y_i assuma apenas um desses valores, pode-se escrever:

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - P)^2 = \frac{N}{N-1} P(1-P).$$

De uma amostra observada s de tamanho n e sendo m o número de elementos da amostra que possuem a característica de interesse, e, do Teorema 3.6, página 38, tem-se para a AAS_S , que um estimador não viciado para P é dado por

$$\hat{P} = \bar{y} = \frac{1}{n} \sum_{i \in s} Y_i = \frac{m}{n}$$

e que

$$V(\hat{P}) = \left(1 - \frac{n}{N}\right) \frac{S^2}{n} = \frac{N-n}{N-1} \frac{PQ}{n},$$

sendo $Q = 1 - P$.

Do Teorema 3.7, página 38, tem-se que um estimador não viciado de S^2 é dado por

$$s^2 = \frac{n}{n-1} \hat{P}\hat{Q},$$

sendo $\hat{Q} = 1 - \hat{P}$.

Assim, tem-se que um estimador não viciado para $V(\hat{P})$ é dado por

$$\hat{V}(\hat{P}) = \left(1 - \frac{n}{N}\right) \frac{\hat{P}\hat{Q}}{n-1}.$$

Um intervalo de confiança aproximado para P é dado por

$$\left(\hat{P} - z_{(\alpha/2)} \sqrt{\left(1 - \frac{n}{N}\right) \frac{\hat{P}\hat{Q}}{n-1}}; \hat{P} + z_{(\alpha/2)} \sqrt{\left(1 - \frac{n}{N}\right) \frac{\hat{P}\hat{Q}}{n-1}} \right).$$

Sendo PQ sempre menor ou igual a $\frac{1}{4}$, tem-se que um intervalo de confiança conservativo para P é dado por

$$\left(\hat{P} - z_{(\alpha/2)} \sqrt{\frac{1-n/N}{4(n-1)}}; \hat{P} + z_{(\alpha/2)} \sqrt{\frac{1-n/N}{4(n-1)}} \right).$$

3.2.3 Comparação entre AAS_C e AAS_S

Considerando a variância dos estimadores como critério de comparação dos planos amostrais, tem-se o *Efeito de Planejamento - EPA*, definido como

$$EPA = \frac{V_{AAS_S}[\bar{y}]}{V_{AAS_C}[\bar{y}]} = \frac{(1 - \frac{n}{N}) \frac{S^2}{n}}{\frac{\sigma^2}{n}} = \frac{N - n}{N - 1}.$$

A estatística \bar{y} , em ambos os planos, é um estimador não viciado para a média populacional μ .

Quando o quociente da expressão acima for um número maior do que um, ou seja, quando $EPA > 1$, tem-se que o plano amostral do numerador é *menos eficiente* (ver Definição 3.17, página 29) que o plano do denominador e, quando este coeficiente for menor do que um, $EPA < 1$, tem-se a situação inversa.

Como na expressão definida para o EPA, o termo final dessa equação é $\frac{N - n}{N - 1}$, e, qualquer que seja o valor de n ($n > 1$), pode-se perceber claramente que esse resultado é sempre menor do que 1 ($EPA < 1$), tem-se que o plano amostral AAS_S é sempre *mais eficiente* do que o plano AAS_C .

3.3 Amostragem Estratificada - AE

Constitui um processo aperfeiçoado de amostragem probabilística, que possibilita a decomposição de um "universo estudado" em "estratos", tendo em vista sua transformação, geralmente em "sub-universos" ou "subpopulações" com menor variabilidade das respectivas características.

A estratificação, no acertado dizer das professoras universitárias, Eva Nick e Sheilah R. de O. Kellner, representa essencialmente um artifício para ganhar em precisão. Realmente, os resultados das pesquisas, assim realizadas, evidentemente que são mais precisos que os obtidos através de amostragem aleatória simples. Trata-se, portanto, de um refinamento técnico-científico, muito embora o custo operacional das referidas investigações, com sua adoção, seja mais elevado (CRUZ, 1978, grifo do autor).

Quando se refere ao custo mais elevado, comparado com a Amostra Aleatória Simples, é devido ao custo operacional da estratificação.

Nessa técnica de amostragem a população é dividida em subpopulações ou estratos, homogêneos internamente, mas heterogêneos entre si, segundo alguma característica conhecida na população em estudo. Em cada um desses estratos são selecionadas amostras em proporções convenientes.

Um dos objetivos da estratificação é a melhoria na precisão dos estimadores da média e do total da população. Se todos os estratos forem homogêneos internamente, no sentido de que o valor das medidas variem pouco de uma unidade para outra, pode-se obter uma estimativa precisa do valor médio de um estrato qualquer, mediante uma

pequena amostra desse estrato. Em seguida, essas estimativas podem ser combinadas para constituírem uma estimativa precisa do conjunto da população.

O plano "Amostragem Estratificada - AE" opera seguindo os procedimentos:

1. Divide-se a população em subpopulações (estratos);
2. Retira-se uma amostra dentro de cada estrato;
3. Em cada amostra, usam-se estimadores adequados para os parâmetros do estrato;
4. Combinando os estimadores de cada estrato obtém-se um estimador para a população.

3.3.1 Notação e Relações Úteis

Considere uma população bem definida e descrita por um sistema de referências

$$U = \{1, 2, \dots, N\}$$

e que exista uma partição U_1, \dots, U_H de U , com

$$U = \bigcup_{h=1}^H U_h \quad \text{e} \quad U_h \cap U_{h'} = \emptyset,$$

para $h \neq h' = 1, \dots, H$.

Também define-se as relações:

- Y : Variável de interesse com $E(Y) = \mu$ e $V(Y) = \sigma^2$;
- N_h : Tamanho do estrato h ;
- $N = \sum_{i=1}^H N_h$: Tamanho da população;
- Y_h : Valores da v. a. de interesse no estrato h , com $E(Y_h) = \mu_h$ e $V(Y_h) = \sigma_h^2$;
- $\tau_h = \sum_{i=1}^{N_h} Y_{hi}$: Total para a variável Y_h no estrato h ;
- $\mu_h = \frac{1}{N_h} \cdot \sum_{i=1}^{N_h} Y_{hi}$: Média para a variável Y_h no estrato h ;
- $S_h^2 = \frac{1}{N_h - 1} \cdot \sum_{i=1}^{N_h} (Y_{hi} - \mu_h)^2$: Variância da variável Y_h no estrato h ;

- $\sigma_h^2 = \frac{1}{N_h} \cdot \sum_{i=1}^{N_h} (Y_{hi} - \mu_h)^2$: Variância da variável Y_h no estrato h ;
- $W_h = \frac{N_h}{N}$: Proporção do estrato h , com $\sum_{h=1}^H W_h = 1$;
- $\tau = \sum_{h=1}^H \tau_h = \sum_{h=1}^H \sum_{i=1}^{N_h} Y_{hi} = \sum_{h=1}^H N_h \mu_h$: Total populacional de Y ;
- $\mu = \bar{Y} = \frac{\tau}{N} = \frac{1}{N} \sum_{h=1}^H \sum_{i=1}^{N_h} Y_{hi} = \frac{1}{N} \sum_{h=1}^H N_h \mu_h = \sum_{h=1}^H W_h \mu_h$: média populacional de Y ;

de modo que a média global é a média ponderada dos estratos. O índice inferior h indica o estrato e o índice i a unidade dentro do estrato.

A diferença primordial entre os termos S_h^2 e σ_h^2 se deve a caracterização da população em finita ou infinita, respectivamente, de acordo com notas encontradas em Cochran (1965), ao definir o item "correção das populações finitas(CPF)".

Conforme pode ser visto em Bolfarine e Bussab (2005), um resultado importante, envolvendo formas quadráticas, estabelece que:

$$\sum_{h=1}^H \sum_{i=1}^{N_h} (Y_{hi} - \mu)^2 = \sum_{h=1}^H \sum_{i=1}^{N_h} (Y_{hi} - \mu_h)^2 + \sum_{h=1}^H N_h (\mu_h - \mu)^2, \quad (3.1)$$

que permite escrever

$$\begin{aligned} \sigma^2 &= \frac{1}{N} \sum_{h=1}^H \sum_{i=1}^{N_h} (Y_{hi} - \mu)^2 \\ &= \frac{1}{N} \left\{ \sum_{h=1}^H \sum_{i=1}^{N_h} (Y_{hi} - \mu_h)^2 + \sum_{h=1}^H N_h (\mu_h - \mu)^2 \right\} \\ &= \frac{1}{N} \left\{ \sum_{h=1}^H N_h \sigma_h^2 + \sum_{h=1}^H N_h (\mu_h - \mu)^2 \right\} \\ &= \sum_{h=1}^H W_h \sigma_h^2 + \sum_{h=1}^H W_h (\mu_h - \mu)^2 \end{aligned} \quad (3.2)$$

ou ainda

$$\sigma^2 = \sigma_d^2 + \sigma_e^2,$$

sendo que

$$\sigma_d^2 = \sum_{h=1}^H W_h \sigma_h^2$$

é a média das variâncias dentro dos estratos e

$$\sigma_e^2 = \sum_{h=1}^H W_h (\mu_h - \mu)^2$$

mede a variação das médias entre os estratos.

Da mesma forma obtém-se

$$S^2 = \sum_{h=1}^H \frac{N_h - 1}{N - 1} S_h^2 + \sum_{h=1}^H \frac{N_h}{N - 1} (\mu_h - \mu)^2.$$

Para estratos relativamente grandes, tem-se

$$S^2 \cong \sigma_d^2 + \sigma_e^2 \cong S_d^2 + \sigma_e^2,$$

onde $S_d^2 = \sum_{h=1}^H W_h S_h^2$.

Quando todos os estratos têm a mesma média, $\mu_h = \mu$, para $h = 1, \dots, H$, a variância populacional σ^2 coincide com σ_d^2 .

Para a média, o total e a variância amostral em cada estrato h , $h = 1, \dots, H$, tem-se:

- $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{hi}$;
- $\hat{\tau}_h = \sum_{i=1}^{n_h} Y_{hi}$ e
- $s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (Y_{hi} - \bar{y}_h)^2$,

com $h = 1, \dots, H$.

Para a amostra toda de tamanho $n = \sum_{h=1}^H n_h$, tem-se:

- $\bar{y} = \frac{1}{n} \sum_{h=1}^H \sum_{i=1}^{n_h} Y_{hi}$;
- $\hat{\tau} = \sum_{h=1}^H \sum_{i=1}^{n_h} Y_{hi}$ e
- $s^2 = \frac{1}{n - 1} \sum_{h=1}^H \sum_{i=1}^{n_h} (Y_{hi} - \bar{y})^2$.

3.3.2 Métodos de Estratificação

Na amostragem estratificada há a necessidade de se identificar como a população pode ser dividida em estratos, de acordo com alguma variável conhecida. Alguns métodos para essa identificação utilizam a própria variável de interesse como variável de estratificação, mas nem sempre isso é possível por "falta de informações" sobre essa variável. Muitas vezes não se tem o conhecimento nem sobre o intervalo de variação da variável de interesse.

A seguir apresentamos alguns métodos de estratificação propostos na literatura. Para isso, consideramos uma população de tamanho N para a qual deseja-se estimar características de interesse com relação a uma variável aleatória Y .

Supor uma variável auxiliar X , correlacionada com Y (podendo, em alguns casos, ser a própria variável Y) e cujos valores são conhecidos para todas as unidades na população.

O problema de estratificação consiste em encontrar os pontos de corte para os estratos, de forma que o estrato h ($h = 2, \dots, H - 1$) é definido como o conjunto de unidades na população, com valores de X pertencentes ao intervalo $(b_{h-1}, b_h]$, sendo que $b_1 < b_2 < \dots < b_{H-1}$ são os limites de cada estrato, o estrato 1 é o conjunto de valores menores ou iguais a b_1 e o estrato H corresponde ao conjunto de observações, cujos valores são maiores que b_{H-1} . A estratificação ótima é aquela que minimiza a variância dos estimadores dos parâmetros de interesse.

Dalenius e Hodges (1959), propõem uma solução para o problema da determinação dos estratos em populações finitas, considerando que a população é representada por uma função densidade de probabilidade $f_X(x)$ conhecida, com média μ e variância σ^2 . Essa abordagem é efetiva somente do ponto de vista teórico, uma vez que na prática, dificilmente é possível definir essa função densidade. O método consiste na construção dos estratos tomando intervalos iguais na função de distribuição acumulada da raiz quadrada de $f_X(x)$. Os pontos de corte $b_1, \dots, b_h, \dots, b_{H-1}$ correspondem à estratificação de menor variância se satisfazem a relação

$$\frac{\sigma_{h-1}^2 + (b_{h-1} - \mu_{h-1})^2}{\sigma_{h-1}} = \frac{\sigma_h^2 + (b_{h-1} - \mu_h)^2}{\sigma_h} \quad (3.3)$$

Na prática é preciso considerar aproximações para (3.3) pois μ_h e σ_h^2 dependem de b_h . No trabalho de Dalenius e Hodges (1959) são apresentadas duas aproximações que requerem a suposição de distribuição Uniforme para X e a escolha arbitrária do número de estratos.

Esse método é utilizado até os dias de hoje, aplicando-se métodos numéricos iterativos para solução. Para populações finitas, os limites ótimos para os estratos podem ser obtidos considerando todas as possíveis divisões da população, associadas com o número de estratos, calculando a variância do estimador de interesse para todas essas divisões e escolhendo a que apresentar a menor variância. No entanto, o número de

possíveis soluções cresce rapidamente com o H e N e, mesmo tendo hoje mais recursos computacionais, esse processo pode ser custoso e demorado. Uma alternativa seria a aplicação de algoritmos iterativos de otimização.

Lavallée e Hidiroglou (1988 apud HORGAN, 2010) propuseram um algoritmo, para o caso de populações assimétricas, que determina os pontos de corte para os estratos, tal que o tamanho da amostra seja mínimo para uma precisão prefixada. Nesse método, inicialmente são considerados pontos de corte de forma que os estratos fiquem com o mesmo tamanho e esses pontos de corte são substituídos iterativamente até que o tamanho mínimo de amostra seja obtido.

Uma outra proposta foi apresentada por Gunning, Horgan e Keogh (2004), na qual a população é dividida em H estratos, considerando quaisquer limites para o intervalo de variação de X . Para aplicação desse método são feitas duas suposições: A primeira, também feita em Dalenius e Hodges (1959), é que a distribuição nos estratos é aproximadamente Uniforme. A segunda, considerada também Dalenius e Hodges (1959), Cochran (1965) e Lavallée e Hidiroglou (1988 apud HORGAN, 2010), é de que os coeficientes de variação (CV) são aproximadamente os mesmos para todos os estratos, isto é:

$$\frac{\sigma_1}{\mu_1} \simeq \dots \simeq \frac{\sigma_H}{\mu_H}$$

Sejam $a = b_0$ o mínimo de X e $ar^H = b_H$ o máximo de X , com $r = (b_H/b_0)^{1/H}$. Os pontos de corte para os estratos h são dados pela progressão geométrica

$$b_h = ar^h, h = 1, \dots, H$$

Nesse artigo o método foi comparado com os métodos de Dalenius e Hodges (1959) e de Lavallée e Hidiroglou (1988 apud HORGAN, 2010) para quatro populações assimétricas e considerando diferentes número de estratos ($H = 3, 4, 5$). A igualdade aproximada dos coeficientes de variação e a precisão dos estimadores, vinculada ao tamanho da amostra, foram usados como critério nessa comparação.

Com relação aos coeficientes de variação, no novo método estes variam muito menos de estrato para estrato do que nos métodos anteriores. Como exemplo, para uma das populações estudadas, no caso de $H = 5$ estratos, temos com o novo método todos os $CV \simeq 0,35$, com o método de Dalenius e Hodges (1959) tem-se os CV variando entre 0,25 e 0,5 e com o método de Lavallée e Hidiroglou (1988 apud HORGAN, 2010) estes variam entre 0,3 e 0,6.

Os resultados mostram também que para a obtenção de estimadores, com uma precisão prefixada, os dois métodos anteriores necessitam de amostras maiores que o novo método, além de que este não requer a escolha arbitrária do número de estratos, como no método de Dalenius e Hodges (1959) e também não requer a criação de limites iniciais como no método de Lavallée e Hidiroglou (1988 apud HORGAN, 2010), portanto a sua implementação é mais fácil.

Pode-se citar ainda alguns trabalhos publicados entre 2000 e 2010 nos quais são apresentados métodos numéricos iterativos de otimização.

Kozak (2004 apud HORGAN, 2010) propõe um algoritmo de busca aleatória tal que o tamanho da amostra, expresso como em Lavallée e Hidiroglou(1988 apud HORGAN, 2010) é minimizado sob as restrições $N_h \geq 2, h = 1, \dots, H - 1$ e $2 \leq n_h \leq N_h, h = 1, \dots, H - 1$; sendo n_h determinado pela alocação ótima de Neyman. O ponto fraco desse método é que ele pode resultar em um tamanho de amostra para o último estrato que excede seu tamanho ($n_H > N_H$).

Keskinturk e Er (2007 apud HORGAN, 2010) sugeriram o uso do algoritmo genético para resolver o problema da estratificação. Eles implementaram o algoritmo para populações reais e simuladas e obtiveram melhores resultados em comparação com o método de Dalenius e Hodges (1959). Khan et al (2008 apud HORGAN, 2010) formularam o problema de estratificação ótima como um problema de programação matemática e desenvolveram a solução utilizando programação dinâmica. Em comparação com o método de Dalenius e Hodges (1959), também mostraram que seu método é mais eficiente.

Em Brito et al. (2010), é apresentado um algoritmo baseado no método do caminho mínimo em grafos e em seu trabalho os autores testaram o algoritmo usando dados reais brasileiros encontrando bons resultados, principalmente com relação à rapidez computacional (tempo de CPU em torno de 3 minutos).

Segundo Horgan (2010), todos esses autores garantem que seus algoritmos resultam em estratificação ótima, minimizando a variância para um dado tamanho de amostra ou minimizando o tamanho da amostra dada a precisão. Todos esses métodos foram aplicados a populações finitas, reais ou simuladas e são computacionalmente intensivos. No entanto, mesmo sabendo que alguns melhoraram alguns aspectos da estratificação, nenhum pode ser considerado melhor que os outros.

Ainda em Horgan (2010), novamente a ideia apresentada em Gunning, Horgan e Keogh (2004) é defendida por ser de fácil implementação e trazer resultados melhores com relação ao coeficiente de variação e com relação à precisão dos estimadores, quando a população estudada é assimétrica (para populações Normais, por exemplo o método não resultará em estratificação ótima).

3.3.3 Definição do plano amostral nos estratos e alocação

Após a definição dos estratos é necessário definir o plano amostral usado dentro de cada estrato. Pode-se utilizar planos distintos para cada estrato. Neste trabalho considerou-se a amostragem aleatória simples em todos os estratos.

Corolário 3.4. *Desprezando as CPF's, de acordo com a justificativa dada na Seção 3.2.2.1, página 37, e, considerando que, dentro de cada estrato, a amostra foi sorteada por um processo AAS_S e que $\hat{\mu}_h = \bar{y}_h$. Então, tem-se para a estimação da média \bar{y}_{es} e*

do total τ_{es} , as seguintes fórmulas:

$$\bar{y}_{es} = \sum_{h=1}^H W_h \bar{y}_h, \quad V(\bar{y}_{es}) = \sum_{h=1}^H W_h^2 \frac{\sigma_h^2}{n_h} \quad e$$

$$\hat{\tau}_{es} = \sum_{h=1}^H N_h \bar{y}_h, \quad V(\hat{\tau}_{es}) = \sum_{h=1}^H N_h^2 \frac{\sigma_h^2}{n_h},$$

com estimadores não viciados dados por

$$\hat{V}(\bar{y}_{es}) = \sum_{h=1}^H W_h^2 \frac{s_h^2}{n_h} \quad e$$

$$\hat{V}(\hat{\tau}_{es}) = \sum_{h=1}^H N_h^2 \frac{s_h^2}{n_h}$$

Após a escolha do plano amostral a distribuição das unidades da amostra nos estratos (alocação) foi estudada. Existem três métodos utilizados para alocação, sendo eles: Alocação Proporcional, Alocação Uniforme e Alocação Ótima de Neyman. Em seguida apresentamos brevemente esses métodos.

3.3.3.1 Alocação Proporcional - AE_{pr}

Com esse procedimento, a amostra aleatória de tamanho n é distribuída proporcionalmente ao tamanho dos estratos, ou seja,

$$n_h = n \frac{N_h}{N} = n W_h, \quad (3.4)$$

A notação AE_{pr} significa "Amostragem Estratificada Proporcional".

Conforme pode ser visto em Bolfarine e Bussab (2005), a AE_{pr} apresenta os seguintes resultados, descritos pelo teorema a seguir:

Teorema 3.8. Na AE_{pr} , o estimador \bar{y}_{es} é igual à média amostral simples \bar{y} , ou seja,

$$\bar{y}_{es} = \bar{y},$$

com

$$V(\bar{y}_{es}) = \sum_{h=1}^H W_h \frac{\sigma_h^2}{n}$$

que é estimada por

$$\hat{V}(\bar{y}_{es}) = \sum_{h=1}^H W_h \frac{s_h^2}{n}.$$

Demonstração. Da média amostral \bar{y}_{es} e da equação (3.4), tem-se que:

$$\begin{aligned}\bar{y}_{es} &= \sum_{h=1}^H W_h \bar{y}_h = \sum_{h=1}^H W_h \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{hi} \\ &= \sum_{h=1}^H W_h \frac{1}{n W_h} \sum_{i=1}^{n_h} Y_{hi} = \frac{1}{n} \sum_{h=1}^H \sum_{i=1}^{n_h} Y_{hi} = \bar{y}.\end{aligned}$$

Tem-se que

$$V(\bar{y}_{es}) = \sum_h W_h^2 \frac{\sigma_h^2}{n_h} \quad (3.5)$$

Mas $\frac{W_h^2}{n_h} = \frac{W_h}{n}$. Então

$$V(\bar{y}_{es}) = \sum_h \frac{W_h}{n} \sigma_h^2 \quad (3.6)$$

Como em cada estrato, s_h^2 é um estimador não viciado para σ_h^2 , então

$$\widehat{V}(\bar{y}_{es}) = \sum_h W_h \frac{s_h^2}{n} = V_{pr}. \quad (3.7)$$

□

3.3.3.2 Alocação Uniforme - AE_{un}

Nessa Alocação, atribui-se o mesmo tamanho de amostra para cada estrato. É utilizada para apresentar estimadores separados para cada estrato.

Para cada um dos H estratos, tem-se:

$$n_h = \frac{n}{H} = k \quad \text{e} \quad f_h = \frac{k}{N_h}.$$

Desse modo tem-se que \bar{y}_{es} é um estimador não viciado para μ .

A variância de \bar{y}_{es} é expressa por:

$$V_{un} = V(\bar{y}_{es}) = \sum_{h=1}^H W_h^2 \left(1 - \frac{k}{N_h}\right) \frac{\sigma_h^2}{k}$$

sendo estimada por

$$\widehat{V}_{un}(\bar{y}_{es}) = \sum_{h=1}^H \left(1 - \frac{k}{N_h}\right) \frac{W_h^2 s_h^2}{k}$$

Observação: Se $\frac{k}{N_h} < 0,05$, não há necessidade desse fator de correção.

3.3.3.3 Alocação Ótima de Neyman - AE_{ot}

Nesse método de alocação são considerados os custos envolvidos no procedimento. Para isso, define-se uma função de custo linear, dada por

$$C = c_0 + \sum_{h=1}^H c_h n_h \quad \text{ou} \quad C' = C - c_0 = \sum_{h=1}^H c_h n_h,$$

sendo que c_0 denota o custo inicial e c_h o custo por unidade observada no estrato h .

Tem-se então que

$$V(\bar{y}_{es}) = \sum_{h=1}^H W_h^2 \frac{\sigma_h^2}{n_h} = V_{es}.$$

O objetivo é minimizar a variância da média amostral para um custo C' fixo ou minimizar o custo C' fixando-se esta variância.

Teorema 3.9. *Na AE com a função de custo linear, tem-se que V_{es} é mínimo para C' fixado ou C' é mínimo para V_{es} , se*

$$n_h = n \frac{W_h \sigma_h / \sqrt{c_h}}{\sum_{h=1}^H \frac{W_h \sigma_h}{\sqrt{c_h}}}, \quad h = 1, \dots, H. \quad (3.8)$$

Demonstração. Minimizar V_{es} com C' fixo ou vice-versa equivale a minimizar o produto $V_{es}C'$.

Tomando

$$V_{es}C' = \left(\sum_{h=1}^H W_h^2 \frac{\sigma_h^2}{n_h} \right) \left(\sum_{h=1}^H c_h n_h \right). \quad (3.9)$$

pode-se utilizar a desigualdade de Cauchy - Schwarz dada por

$$\left(\sum a_h^2 \right) \left(\sum b_h^2 \right) \geq \left(\sum a_h b_h \right)^2, \quad (3.10)$$

A igualdade em 3.10 ocorre se $\frac{b_h}{a_h}$ é constante para $h = 1, \dots, H$.

Então se

$$\frac{\sqrt{c_h n_h}}{W_h \sigma_h / \sqrt{n_h}} = \frac{n_h \sqrt{c_h}}{W_h \sigma_h} = k \quad (3.11)$$

para $h = 1, \dots, H$, o produto $V_{es}C'$ é mínimo se

$$n_h = k \frac{W_h \sigma_h}{\sqrt{c_h}}, \quad (3.12)$$

com $h = 1, \dots, H$.

Fazendo

$$\sum n_h = k \sum_h \frac{W_h \sigma_h}{\sqrt{c_h}} = n \quad (3.13)$$

tem-se

$$k = \frac{n}{\sum W_h \frac{\sigma_h}{\sqrt{c_h}}}. \quad (3.14)$$

Substituindo-se a equação 3.14 em 3.13, obtém-se o resultado esperado. \square

Portanto, a alocação ótima é diretamente proporcional a $N_h \sigma_h$ e inversamente proporcional a $\sqrt{c_h}$.

Corolário 3.5. Para C' fixo ou a V_{es} fixa, o tamanho ótimo da amostra é dado, respectivamente, por:

$$\begin{aligned} \text{i. } C' \text{ fixo} &\implies n = C' \frac{\sum_{h=1}^H \frac{N_h \sigma_h}{\sqrt{c_h}}}{\sum_{h=1}^H N_h \sigma_h \sqrt{c_h}}; \\ \text{ii. } V_{es} \text{ fixa} &\implies n = \frac{1}{V_{es}} \left(\sum_{h=1}^H W_h \sigma_h \sqrt{c_h} \right) \left(\sum_{h=1}^H \frac{W_h \sigma_h}{\sqrt{c_h}} \right), \text{ sendo } W_h = \frac{N_h}{N}. \end{aligned}$$

3.3.4 Efeito do Planejamento

Quando se utiliza a amostragem estratificada, geralmente tem-se uma menor variância para os valores médio e total estimados, em comparação àquela variância obtida quando se utiliza a AAS para o cálculo desses mesmos valores.

Porém, também não é verdade, que uma amostragem estratificada dê sempre uma menor variância em comparação com uma amostra obtida pela AAS. Se os valores de n_h (n° de unidades da amostra no estrato h) estiverem longe do ótimo, ou seja, mal dimensionadas, a AE pode ter variância maior.

Far-se-á nesta seção uma comparação entre a AAS e AE com alocações proporcional e ótima, cujas variâncias são dadas, respectivamente, por

$$V_S = V_{AAS_S}(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{S^2}{n}, \quad (3.15)$$

$$\text{com } S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \mu)^2;$$

$$V_{pr}(\bar{y}) = \sum_{h=1}^H W_h \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n}, \quad (3.16)$$

$$\text{com } S_h^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (Y_{hi} - \mu_h)^2; \text{ e}$$

$$V_{ot}(\bar{y}) = \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n} \quad (3.17)$$

cuja relação entre elas pode ser apresentada pelo teorema seguinte.

Teorema 3.10. *Desprezando-se os termos em $\frac{n_h}{N_h}$, tem-se com relação a AAS_S , que*

$$V_{ot} \leq V_{pr} \leq V_S$$

onde a alocação ótima é para n fixo e com $n_h \propto N_h S_h$.

Demonstração. Desprezando-se as CPF das variâncias nas equações (3.15), (3.16) e (3.17), conforme subseção 3.2.2.1 na página 37, pode-se escrever:

$$V_S(\bar{y}) = \frac{S^2}{n}; \quad V_{pr}(\bar{y}) = \sum_{h=1}^H W_h \frac{S_h^2}{n} \quad e \quad V_{ot}(\bar{y}) = \sum_{h=1}^H W_h^2 \frac{S_h^2}{n}.$$

De acordo com a identidade algébrica-padrão (equação 3.1, página 43) para a análise de variância das populações estratificadas, tem-se

$$\begin{aligned} (N-1) S^2 &= \sum_{h=1}^H \sum_{i=1}^{N_h} (Y_{hi} - \mu)^2 \\ &= \sum_{h=1}^H \sum_{i=1}^{N_h} (Y_{hi} - \mu_h)^2 + \sum_{h=1}^H N_h (\mu_h - \mu)^2 \\ &= \sum_{h=1}^H (N_h - 1) S_h^2 + \sum_{h=1}^H N_h (\mu_h - \mu)^2. \end{aligned} \quad (3.18)$$

Uma vez que os termos em $\frac{1}{N_h}$ são desprezíveis, tem-se

$$NS^2 = \sum_{h=1}^H N_h S_h^2 + \sum_{h=1}^H N_h (\mu_h - \mu)^2 \quad (3.19)$$

Logo, tem-se que

$$\begin{aligned} V_S &= \frac{\sum_{h=1}^H N_h S_h^2}{nN} + \frac{\sum_{h=1}^H N_h (\mu_h - \mu)^2}{nN} \\ V_S &= V_{pr} + \frac{\sum_{h=1}^H N_h (\mu_h - \mu)^2}{nN} \end{aligned} \quad (3.20)$$

Como o termo $\frac{\sum_{h=1}^H N_h (\mu_h - \mu)^2}{nN}$ é sempre não negativo e V_S sendo V_{pr} mais um fator, pode-se então concluir que

$$V_{pr} < V_S.$$

Partindo-se do Teorema 3.10, tem-se que $V_{pr} \geq V_{ot}$. A diferença entre essas duas

variâncias é dada por

$$\begin{aligned}
 V_{pr} - V_{ot} &= \sum_{h=1}^H W_h \frac{S_h^2}{n} - \sum_{h=1}^H W_h^2 \frac{S_h^2}{n} \\
 &= \sum_{h=1}^H \frac{N_h}{N} \frac{S_h^2}{n} - \sum_{h=1}^H \frac{N_h^2}{N^2} \frac{S_h^2}{n} \\
 &= \frac{1}{nN} \left(\sum_{h=1}^H N_h S_h^2 - \frac{\sum_{h=1}^H (N_h S_h)^2}{N} \right) \\
 &= \frac{1}{nN} \sum_{h=1}^H N_h (S_h - \bar{S})^2, \tag{3.21}
 \end{aligned}$$

onde $\bar{S} = \frac{\sum_{h=1}^H N_h S_h}{N}$.

Combinando as expressões (3.20) com (3.21), tem-se que

$$V_S = V_{ot} + \underbrace{\frac{\sum_{h=1}^H N_h (S_h - \bar{S})^2}{n N}}_A + \underbrace{\frac{\sum_{h=1}^H N_h (\mu_h - \mu)^2}{n N}}_B. \tag{3.22}$$

Na equação (3.22) pode-se verificar os componentes que influenciam na diminuição da variância quando se passa da AAS_S para AE com alocação ótima. O termo B, provém da eliminação das diferenças entre os valores médios dos estratos, enquanto que o termo A, é proveniente da eliminação do efeito das diferenças entre os desvios-padrões dos estratos, representando, assim, a diferença de variância entre a alocação proporcional e ótima.

No caso das CPF's não serem desprezíveis, a análise dos referidos dados nos conduz a equação

$$V_S = V_{pr} + \frac{N - n}{nN(N - 1)} \left[\sum_{h=1}^H N_h (\mu_h - \mu)^2 - \frac{1}{N} \sum_{h=1}^H (N - N_h) S_h^2 \right]$$

Conclui-se então, que a AE com alocação proporcional dará uma variância maior que AAS_S se

$$\sum_{h=1}^H N_h (\mu_h - \mu)^2 < \frac{1}{N} \sum_{h=1}^H (N - N_h) S_h^2.$$

□

3.3.5 Estimação das Proporções

Suponha que se queira conhecer a ocorrência de determinada característica na população. Associando-se a cada elemento da população os valores 1 e 0, conforme esta

unidade apresente a característica desejada ou não, e definindo-se o vetor de coordenadas (h, i) , então a quantidade de interesse associada ao i -ésimo elemento no h -ésimo estrato pode ser representado por

$$Y_{hi} = \begin{cases} 1, & \text{se o elemento } (h, i) \text{ possui a característica} \\ 0, & \text{caso contrário.} \end{cases}$$

Sendo $\tau_h = \sum_{i=1}^{N_h} Y_{hi}$, o número de elementos na amostra que possui a característica de interesse no estrato h , tem-se que

$$P_h = \frac{\tau_h}{N_h} = \mu_h$$

é a proporção desses elementos no estrato h , $h = 1, \dots, H$, sendo que para a amostra toda essa proporção pode ser escrita como

$$P = \sum_{h=1}^H W_h P_h.$$

Dada uma amostra de tamanho n_h selecionada segundo a AAS_C no estrato h pode-se definir para P o estimador

$$\hat{P}_{es} = \sum_{h=1}^H W_h \hat{P}_h,$$

onde $\hat{P}_h = \bar{y}_h$.

Fazendo $\hat{P}_{es} = \bar{y}_{es}$, tem-se

$$\sigma_h^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (Y_{hi} - P_h)^2 = P_h (1 - P_h).$$

Com relação à AE com reposição, tem-se que

$$\hat{P}_{es} = \bar{y}_{es} = \sum_{h=1}^H W_h \hat{P}_h$$

é um estimador não viciado de P com

$$V(\hat{P}_{es}) = \sum_{h=1}^H W_h^2 \frac{P_h Q_h}{n_h},$$

onde $Q_h = 1 - P_h$, $h = 1, \dots, H$.

Conseqüentemente, um estimador não viciado para essa variância, com relação à AE com reposição é dada por

$$\hat{V}(\hat{P}_{es}) = \sum_{h=1}^H W_h^2 \frac{\hat{P}_h \hat{Q}_h}{n_h - 1}.$$

3.3.6 Intervalos de Confiança

Conforme o tamanho da amostra aumenta, as distribuições de \bar{y}_{es} e $\hat{\tau}_{es}$ vão se aproximando da distribuição normal, segundo o TCL. Assim, para n_h e N_h suficientemente grandes, tem-se que:

$$\frac{\bar{y}_{es} - \mu}{\sqrt{\sum_{h=1}^H \frac{W_h^2 \sigma_h^2}{n_h}}} \cong N(0, 1) \quad (3.23)$$

e que,

$$\frac{\hat{\tau}_{es} - \tau}{\sqrt{\sum_{h=1}^N \frac{N_h^2 \sigma_h^2}{n_h}}} \cong N(0, 1) \quad (3.24)$$

Como σ_h^2 não é conhecido nas equações acima, utiliza-se nesse caso o seu estimador não viciado s_h^2 . Assim, um intervalo de confiança para μ com coeficiente de confiança igual a $1 - \alpha$ é dado por

$$IC(\mu_{es}) = \bar{y}_{es} \pm z_{(\alpha/2)} \sqrt{\sum_{h=1}^H W_h^2 \frac{s_h^2}{n_h}}. \quad (3.25)$$

O intervalo de confiança para τ é obtido de forma semelhante.

Para o caso das proporções e considerando-se uma distribuição normal, um intervalo de confiança para P , com coeficiente de confiança $1 - \alpha$ é dado por

$$IC(P_{es}) = \hat{P}_{es} \pm z_{(\alpha/2)} \sqrt{\sum_{h=1}^H W_h^2 \frac{\hat{P}_h \hat{Q}_h}{n_h - 1}}.$$

3.3.7 Determinação do Tamanho da Amostra

Além das equações descritas pelo Teorema 3.9, página 50 e pelo Corolário 3.5, página 51, dentro do estrato h e considerando-se a função de custo linear $C = c_0 + \sum_{h=1}^H c_h n_h$, a alocação ótima de Neyman é dada por

$$n_h = n \frac{N_h \sqrt{P_h Q_h / c_h}}{\sum_{h=1}^H N_h \sqrt{P_h Q_h / c_h}}. \quad (3.26)$$

Observação 3.1. Não se dispondo de informação preliminar sobre P_h , através de amostras pilotos ou pesquisas anteriores, pode-se substituir $P_h Q_h$ na expressão 3.26 por $\frac{1}{4}$, por ser um limite superior, levando à alocação proporcional.

Para os casos gerais, ou seja, uma alocação qualquer, fixada a variância V , o valor de n poderá ser obtido por

$$n = \frac{\sum_{h=1}^H \frac{W_h^2 s_h^2}{w_h}}{V + \frac{1}{N} \sum_{h=1}^H W_h s_h^2}, \quad (3.27)$$

cuja variância prevista para \bar{y}_{es} é obtida pela fórmula

$$V = \frac{1}{n} \sum_{h=1}^H \frac{W_h^2 s_h^2}{w_h} - \frac{1}{N} \sum_{h=1}^H W_h s_h^2.$$

Se as CPF's forem desprezadas, tem-se, como uma primeira aproximação,

$$n_0 = \frac{1}{V} \sum_{h=1}^H \frac{W_h^2 s_h^2}{w_h}.$$

Se $\frac{n_0}{N}$ não for desprezível, podemos calcular n pela fórmula

$$n = \frac{n_0}{1 + \frac{1}{NV} \sum_{h=1}^H W_h s_h^2}.$$

Nos casos particulares, as fórmulas acima tomam várias formas, que podem ser mais cômodas para os cálculos, onde algumas destas são:

i. Alocação ótima presumível (sendo n fixado) com w_h equivalente a $W_h s_h$.

$$n = \frac{(\sum_{h=1}^H W_h s_h)^2}{V + \frac{1}{N} \sum_{h=1}^H W_h s_h^2}$$

ii. Alocação proporcional com $w_h = W_h = \frac{N_h}{N}$.

$$n_0 = \frac{\sum_{h=1}^H W_h s_h^2}{V} \quad \text{e} \quad n = \frac{n_0}{1 + \frac{n_0}{N}}$$

Para as referidas fórmulas cita-se a bibliografia Cochran(1965), e sua seção 5.9.

3.4 Amostragem por Conglomerados - AC

A amostragem por conglomerados é utilizada quando a população é muito dispersa, inviabilizando a organização de um rol com todos os seus elementos. Representa o tipo mais prático e também o mais econômico da amostragem probabilística. A população é dividida em subpopulações ou conglomerados, os quais são selecionados segundo os critérios da AAS. Os elementos dentro dos conglomerados devem apresentar características semelhantes.

Como regra geral, o número de elementos em um conglomerado deve ser pequeno em relação ao tamanho da população, e o número de conglomerados, razoavelmente grande. No caso do total populacional ser representado por um município, pode-se considerar como conglomerados os seus bairros, as ruas ou os quarteiros. Todos os elementos dos conglomerados sorteados serão amostrados.

Como desvantagem, pode-se citar o fato de que as unidades, dentro de um mesmo conglomerado, tendem a ter valores bem próximos em relação às variáveis que estão sendo pesquisadas, acarretando com isso os planos amostrais menos eficientes, além de apresentar maiores dificuldades nas análises estatísticas, quando comparadas com os modelos da AAS ou mesmo AE.

3.5 Amostragem em Múltiplos Estágios - AME

Nesse método a população também é dividida em subpopulações, assim como nos métodos da AE ou AC. O presente método admite, nas diversas etapas de sua execução, o emprego (conforme a natureza do fenômeno estudado) dos dois primeiros tipos de amostragem probabilística, ou seja, a amostragem aleatória simples e/ou a amostragem estratificada, contribuindo, assim, para torná-lo mais flexível. Esse tipo de amostragem é muito utilizada em populações humanas ou animal, onde considerando-se uma cidade como sendo o conjunto do total populacional, em um primeiro estágio, algumas subunidades são selecionadas usando-se os critérios da AAS, como por exemplo os bairros, que representam as UPA's (unidade primária de amostragem). Num segundo estágio, onde tem-se as USA's (unidade secundária de amostragem), uma amostra é selecionada de cada subpopulação que foi selecionada no primeiro estágio, onde pode-se ter no caso os quarteirões. Num terceiro estágio, pode-se ter os domicílios como as unidades elementares de interesse. Nesse caso, tem-se uma amostragem por conglomerados em três estágios.

3.6 Amostragem Sistemática - AS

Nesse método de amostragem é exigida uma listagem dos indivíduos da população. Uma das vantagens que esse método apresenta em relação à AAS é a facilidade na seleção de uma amostra, a qual pode inclusive acontecer diretamente no campo. A amostra se distribui da maneira mais uniforme na população. Este fato torna, às vezes, a amostragem sistemática consideravelmente mais precisa do que a amostragem aleatória simples.

Para a operação dessa técnica de amostragem, considere N o tamanho da população e n o tamanho da amostra, calculando-se em seguida o fator de sistematização k pela expressão $k = \frac{N}{n}$, sendo k um número inteiro mais próximo do resultado obtido por esse cálculo. Na sequência, utilizando os critérios da AAS, sorteia-se um número qualquer entre 1 e k , digamos m , sendo $0 < m < k$. Esse número m será o primeiro elemento da amostra, onde a partir daí, obtêm-se os demais elementos dessa amostra de forma sistemática, acrescentando sempre ao elemento anterior o valor de k . Assim, a amostra será formada pelos elementos: $(m, m + k, m + 2k, \dots)$, e assim sucessivamente até completar toda amostra, cujo número de elementos é previamente estipulado.

Suponha, por exemplo, que uma empresa de telefonia fixa deseja conhecer o grau de satisfação de seus usuários com os serviços prestados, onde o número de assinantes é da ordem de 50.000. Deseja-se selecionar uma amostra aleatória de 1.000 destes assinantes com o intuito de obter uma avaliação sobre os serviços.

Calculando-se o valor de k , $k = \frac{50.000}{1.000} = 50$, sorteia-se, através dos critérios da AAS, um número qualquer m , dentre os 50 primeiros elementos dessa população, digamos $m = 30$, e a partir daí forma-se sistematicamente a amostra, somando o valor de k ao elemento anterior, cuja amostra será assim formada pelos seguintes elementos da população: $s = (30, 80, 130, 180, \dots)$.

3.7 Exemplo de Aplicação

Exemplo 3.1. Nessa seção será apresentado um exemplo que compara os resultados de uma amostra analisada através dos critérios da amostra aleatória simples sem reposição e em seguida comparada com os resultados obtidos através de uma análise segundo os critérios da amostragem estratificada.

Considere inicialmente os dados brutos da tabela constante do anexo I, correspondente a área, em hectares, cultivadas com a cultura do milho no município de Lavras - MG na safra de 1988. Nesse trabalho foram levantadas informações em 194 propriedades daquele município, conforme pode ser visto em Muniz e Abreu (1999).

Admite-se que tenha sido aplicado, à rigor, os critérios de seleção para uma amostra aleatória simples, sem reposição, para o sorteio de 25 propriedades constantes daquele universo, obtendo-se os seguintes elementos (Y_i), no que se refere às áreas cultivadas:

Tabela 3.1: Valores dos Y_i (Áreas Cultivadas)

1,4	6,1	30,0	11,4	62,0	1,0	3,6	1,5	8,2	5,4	53,0	90,0	35,0
8,8	1,8	1,3	4,9	9,0	20,8	17,3	8,2	4,2	1,2	5,9	8,3	

Tem-se os seguintes Cálculos:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n Y_i = 16,01,$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{y})^2 = 494,14.$$

Cálculo da fração amostral $\frac{n}{N} = \frac{25}{194} = 0,13$. Como $\frac{n}{N} > 0,05$ a população é considerada finita. Assim:

$$\widehat{V}(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{s^2}{n} = 17,2178.$$

Cálculo do intervalo de confiança para \bar{y} . Considerando um grau de liberdade $GL = 24$ e um coeficiente de confiança $1 - \alpha = 0,95$, tem-se pela tabela de distribuição de Student, que o valor de $t_{(\alpha/2)} = 2,064$. Assim:

$$IC(\mu) = \bar{y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 16,01 \pm 9,18 = (6,83; 25,19).$$

Admite-se agora que a população do universo de estudo foi estratificada, sendo sorteada em cada estrato as seguintes amostras, conforme Tabela 3.2.

Tabela 3.2: Distribuição dos elementos da amostra por estrato

Est.	Área(ha)	N_h	n_h	Amostras						\bar{y}
1	< 4,0	60	5	1,6	1,0	3,7	2,4	1,8		2,10
2	[4,0; 10,0)	49	6	8,9	7,3	8,2	4,5	5,9	7,6	7,07
3	[10,0; 20,0)	35	6	12,2	17,8	15,0	11,4	14,0	14,6	14,17
4	[20,0; 40,0)	30	4	35,3	29,7	27,0	22,0			28,50
5	$\geq 40,0$	20	4	82,0	62,0	75,0	54,0			68,25
Total		194	25							

Os cálculos para o estrato 1, serão apresentados abaixo e para os demais estratos procede-se de forma semelhante, ou seja:

$$\bar{y}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} Y_{1i} = 2,10,$$

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (Y_{1i} - \bar{y}_1)^2 = 1,05,$$

$$\widehat{V}(\bar{y}_1) = \left(1 - \frac{n_1}{N_1}\right) \frac{s_1^2}{n_1} = 0,1925.$$

Cálculo do intervalo de confiança para \bar{y}_1 . Considerando um grau de liberdade $GL = 4$ e um coeficiente de confiança $1 - \alpha = 0,95$, tem-se pela tabela de distribuição de Student, que o valor de $z_{(\alpha/2)} = 2,776$. Assim:

$$IC(\mu_1) = \bar{y}_1 \pm t_{\alpha/2} \frac{s_1}{\sqrt{n_1}} = 2,10 \pm 1,27 = (0,83; 3,37).$$

O resumo dos cálculos em todos os estratos é mostrado na Tabela 3.3. Os referidos cálculos foram efetuados em planilhas do Microsoft Excel.

Tabela 3.3: Resumo dos cálculos por estrato

Est.	Área(ha)	N_h	n_h	\bar{y}_h	s_h^2	$\widehat{V}(\bar{y}_h)$	$t_{0,025}$	IC p/ \bar{y}_h
1	< 4,0	60	5	2,10	1,05	0,1925	2,776	(0,83; 3,37)
2	[4, 0; 10, 0)	49	6	7,07	2,59	0,3788	2,571	(5,38; 8,76)
3	[10, 0; 20, 0)	35	6	14,17	5,13	0,7085	2,571	(11,79; 16,55)
4	[20, 0; 40, 0)	30	4	28,50	30,73	6,6584	3,182	(19,68; 37,32)
5	\geq 40,0	20	4	68,25	158,92	31,7820	3,182	(48,19; 88,31)
Total		194	25					

Por esta tabela se pode notar a grande variabilidade que existe entre os estratos. Pelo cálculo de s_h^2 essa variabilidade do estrato 5 em relação ao estrato 1 é da ordem de 151 vezes maior.

Para a amostra toda, têm-se os seguintes cálculos:

$$\bar{y}_{es} = \frac{1}{N} \sum_{h=1}^H N_h \bar{y}_h = 16,42,$$

$$\widehat{V}(\bar{y}_{es}) = \sum_{h=1}^H W_h^2 \widehat{V}(\bar{y}_h) = 0,5616.$$

Cálculo do intervalo de confiança para \bar{y}_{es} . Considerando um grau de liberdade $GL = 7$ e um coeficiente de confiança $1 - \alpha = 0,95$ tem-se pela tabela de distribuição de Student, que o valor de $z_{(\alpha/2)} = 2,064$. Assim:

$$IC(\mu) = \bar{y}_{es} \pm t_{\alpha/2} \sqrt{V(\bar{y}_{es})} = 16,42 \pm 1,77 = (14,65; 18,19).$$

Dos resultados obtidos nesse exemplo, pode-se notar uma maior precisão nos resultados obtidos quando se utiliza a amostragem estratificada em relação a amostra dimensionada pela amostragem aleatória simples. Enquanto que na AAS_S um intervalo de confiança para a média ficou em (6,83; 25,19), na AE esse mesmo intervalo ficou reduzido para (14,65; 18,19), reduzindo, assim, um erro de estimação da ordem de 80% quando utilizada esta última técnica.

4 Aplicação com dados reais

4.1 O Projeto: População Canina

4.1.1 Introdução

Em grande parte do mundo e também nos municípios brasileiros as populações de cães representam um problema de saúde pública. Enfrenta-se o risco de transmissão de zoonoses, tais como a raiva, a leishmaniose visceral e o complexo larva migrans visceral/toxocaríose, conforme descrito em Matos (2002). Também há o risco de agressão por mordidas, principalmente em crianças. Conforme pode ser visto em Soto et al. (2006), em 1998 foram notificados 18.000 casos de mordidas desses animais em seres humanos, apenas na cidade de São Paulo, acarretando com isso despesas com atendimento médico, faltas ao trabalho, além de outros prejuízos indiretos. Apesar das deficiências nas avaliações quantitativas, são conhecidos também registros de casos de acidentes de trânsito provocados pela presença desses animais soltos nas vias públicas das cidades, como também nas estradas, conforme descrito em WHO/WSPA (1992 apud MATOS et al. 2002).

As ações primárias de controle de zoonoses em cães, desenvolvidas nos municípios, estiveram pautadas na captura e na eutanásia desses animais, além da aplicação de leis e regulamentos para aqueles proprietários mais omissos no controle de seus animais de estimação. O recolhimento de cães nas vias públicas e domicílios como prática de controle das zoonoses é uma das medidas mais antigas e tradicionais. De acordo com Soto et al. (2006), somente na cidade de São Paulo, no período de 1997 a 2002, foram eliminados em média 25 a 30 mil cães por ano. No entanto, essa prática gera despesas aos cofres públicos, não resolvendo o problema da população elevada, onde qualquer redução no tamanho dessa população por aumento na mortalidade de animais, é rapidamente compensada pelo aumento na reprodução e na taxa de sobrevivência.

Segundo a OMS (Organização Mundial da Saúde), não existe prova de que a eliminação de cães tenha gerado impacto significativo na densidade das populações caninas e mesmo na propagação da raiva. A renovação das populações caninas é muito rápida e a taxa de sobrevivência delas se sobrepõe à taxa de eliminação. Os métodos aceitáveis para o controle da população canina são baseados em restrição de movimentos, controle

de habitat e regulação da produção. Conhecer a estrutura e o tamanho dessa população em uma localidade é importante para o planejamento de programas de controle dessa população e, conseqüentemente de suas zoonoses. Vários estudos foram realizados no Brasil e no exterior para estimar parâmetros relacionados a essas populações, alguns deles com base em procedimentos amostrais e outros com base em censos. Ainda que os censos permitam uma avaliação mais precisa de parâmetros da população, nem sempre é possível realizá-los devido a dificuldades financeiras e disponibilidade de pessoal técnico capacitado.

Para o planejamento de ações preventivas e o controle das zoonoses, também é importante conhecer os parâmetros específicos da população de cães de rua, cuja existência apresenta relação direta com o meio ambiente e os aspectos culturais da comunidade local. Portanto, além do conhecimento sobre a densidade populacional, é importante fazer uma classificação dos animais de acordo com os distintos níveis de cuidados que recebem.

De acordo com a dependência e a restrição, a OMS classifica os cães em: supervisionados ou controlados, de família, de vizinhança ou comunitários e selvagens, conforme pode ser visto em WHO/WSPA (1992 apud MATOS et al. 2002). Para as duas primeiras classes os levantamentos domiciliares permitem aferições precisas, mas para a parcela de animais que vivem soltos pelas ruas a avaliação é prejudicada. As técnicas de captura e recaptura disponíveis para o dimensionamento, tem como pré-requisito a marcação e a permanência em campo por vários dias a fim de se obter os dados necessários à estimação final do tamanho da população desses animais. Algumas dessas técnicas requer a manipulação dos animais implicando na participação de pessoas com habilidade e especialização para contê-los e marcá-los, o que dificulta e onera os procedimentos de campo. Além disso, sua aplicação não permite identificar os animais segundo a classificação definida pela OMS.

Pesquisadores do Instituto Pasteur e da Superintendência de Controle de Endemias de São Paulo desenvolveram uma técnica para dimensionar e classificar populações caninas, segundo grau de dependência e restrição. A técnica consiste na realização de inquérito domiciliar para identificação de animais com proprietários e o seu grau de restrição, e contagem dos cães que circulam pelas ruas. Essa técnica, desenvolvida para coletar informações de difícil acesso, permite obtê-las com rapidez, simplicidade e baixo custo, uma vez que pode ser realizada por pessoal não especializado, treinado por curto período de tempo e recrutado entre a população local, evitando assim, despesas com deslocamentos e pagamento de mão-de-obra especializada. Nesse método, não existe a manipulação de animais por parte dos pesquisadores, significando com isso um menor risco de acidentes e economia com equipamentos de proteção, conforme descrito em Matos et al. (2002).

4.1.2 A técnica Pasteur São Paulo

A técnica Pasteur São Paulo foi implementada de acordo com o procedimento apresentado por Matos et al. (2002), sendo, portanto, considerada a seguinte classificação para a população canina de Rio Claro - SP, objeto de estudo, em quatro classes, a saber:

- i. c_1 : Cães domiciliados: animais que recebem alimento, abrigo e cuidados gerais do proprietário e são totalmente restritos (não saem às ruas sem os proprietários);
- ii. c_2 : Cães semi-domiciliados: animais que recebem alimento e abrigo, mas os cuidados gerais são menos rigorosos, saindo às ruas sem acompanhamento e restrições;
- iii. c_3 : Cães de vizinhança: animais que permanecem próximos as residências ou estabelecimentos comerciais, recebendo alimentos, mas sem um proprietário definido;
- iv. c_4 : Cães sem proprietários.

Esses dois últimos grupos estão juntos na classificação da OMS, chamados de cães de vizinhança. Na técnica Pasteur São Paulo são considerados separados por se referirem a categorias para as quais podem ser indicadas medidas distintas de controle. Os cães classificados como selvagens pela OMS são excluídos, pois dificilmente são encontrados no ambiente urbano.

4.1.3 Entrevistas domiciliares

Um esboço do questionário contendo as questões para a caracterização da população canina com relação aos graus de dependência e restrição encontra-se no Anexo B.

A aplicação dos questionários foi realizada considerando-se uma amostra dos domicílios de Rio Claro. O plano amostral foi a Amostragem Estratificada, conforme descrito na Seção 3.3, página 41. Em cada domicílio, após a aplicação do questionário e classificação dos cães nas categorias 1 ou 2, os cães da categoria 2 receberam uma coleira para serem identificados por ocasião da contagem dos cães de rua. As pessoas que relataram cuidar de algum cão da vizinha foram orientadas a colocar coleiras nesses animais de uma segunda cor, assim que eles aparecerem em busca de alimentos, desde que alguém ainda não o tivesse feito.

Essa etapa foi realizada por estudantes do ensino superior e voluntários, que residem na área de estudo para evitar custos elevados com transportes e alimentação. Essas pessoas receberam treinamento que consistiu da apresentação de aspectos relacionados à Saúde Pública, objetos e instrumentos da coleta de dados da pesquisa.

4.1.4 Contagem dos cães fora dos domicílios

No dia seguinte ao inquérito domiciliar do primeiro estrato visitado (estrato 3), os pesquisadores foram distribuídos estrategicamente em toda a área, em locais que

permitissem uma visualização das cores das coleiras dos cães que circulavam pelas ruas. A distância máxima de observação era de 50 metros à direita e à esquerda. A classe de pertinência dos cães era identificada pela cor das coleiras que possuíam, sendo que a sua ausência indicaria que os cães pertenciam a classe 4.

Os pontos de observação foram definidos previamente e para que não houvesse sobreposição de áreas e duplicidade de registros, os pesquisadores foram orientados para não se deslocarem por mais de 2 metros do seu ponto de observação.

O tempo de observação foi de 30 segundos de forma que o pesquisador pudesse registrar em sua planilha o que viu, sem a necessidade de memorizar a informação. Essas contagens foram repetidas a cada 15 minutos, durante uma hora, possibilitando o registro do maior número de cães. O início e término de cada período de 30 segundos de observação foi marcado por relógios do próprio pesquisador, os quais foram acertados antecipadamente pelo relógio da coordenadora do projeto, de forma a garantir a simultaneidade dessas observações.

Devido a escassez de mão de obra para essa contagem a mesma ficou prejudicada para esse estrato, sendo inclusive suspensa nos estratos seguintes, acarretando com isso a falta de informações para os cães das categorias 3 e 4.

Nos condomínios fechados não foi possível fazer inquérito domiciliar, pela dificuldade em obter autorização dos síndicos. O único condomínio visitado foi o Residencial Florença.

4.2 Resultados e Discussões

4.2.1 Campo da Pesquisa

O presente estudo foi realizado no município de Rio Claro, estado de São Paulo, onde há presença de animais da espécie canina circulando livremente em meio a população, principalmente nas áreas com baixo poder aquisitivo, razão pela qual é imprescindível o conhecimento da população desses animais, suas relações com a população humana e os riscos à saúde pública.

O município de Rio Claro está localizado na região Centro-Leste do estado, com uma população urbana de 181.766 habitantes (censo de 2010) sendo as suas atividades econômicas oriundas principalmente da Indústria e Agricultura. No setor industrial, merece destaque as fábricas de cerâmicas e diversidade de vários produtos industrializados, enquanto que na agricultura, destaca-se a produção de cana-de-açúcar.

Para a estratificação da população da zona urbana, considerou-se uma divisão feita pela Prefeitura Municipal de Rio Claro, distribuída em cinco regiões denominadas de "Unidades Regionais de Planejamento" (URP's), que, segundo o SEPLADEMA (Secretaria Municipal de Planejamento, Desenvolvimento e Meio Ambiente) esta divisão se deu com base nas condições sócio-econômicas da população de cada área, entre outros

critérios. O IBGE informou o número de domicílios por setor censitário e procedeu-se o cálculo do número de domicílios por URP utilizando mapas detalhados da área urbana. Para este trabalho, estas URP's serão designadas por estratos. As comunidades entrevistadas encontram-se relacionadas a seguir, segundo a URP (ou estrato) a que pertencem.

- i. URP_1 : Alto do Santana, Bairro da Saúde, Cidade Jardim, Jd. Claret, Jd. Araucária, Jd. Boa Vista, Jd. do Trevo, Jd. Itapoã, Jd. Kennedy, Jd. Mirassol, Jd. Olinda, Jd. Santa Eliza, Pq. São Paulo, Pq. Universitário, Santana, Vila Operária, Vila Santa Cruz e Zona Central.
- ii. URP_2 : Boa Esperança, Cherveson, Hipódromo, Jd. Floridiana, Jd. Progresso, Jd. Santa Maria e Pq. das Indústrias.
- iii. URP_3 : Mãe Preta, Rec. Águas Claras, Res. Florença e Vila Industrial.
- iv. URP_4 : Bela Vista, Cidade Nova, Indaiá, Jd. Vilage, Jd. América, Jd. Bandeirantes, Jd. Conduta, Jd. dos Ipês, Jd. Flórida, Jd. Paulista, Jd. São José, Jd. Bela Vista, Orestes A. Giovani, São Miguel, Vila Alemã e Vila Nova.
- v. URP_5 : Jd. Inocoop, Jd. Brasília, Jd. Guanabara, Jd. Guanabara II, Jd. Nova Rio Claro, Jd. Bom Sucesso, Jd. Centenário, Jd. Novo, Jd. Novo II, Jd. Novo Wenzel, R. dos Bosques, Res. Palmeiras.

Para a alocação do número de domicílios por estrato utilizou-se a alocação ótima de Neyman, conforme descrito pela equação 3.26, página 55, efetuando-se várias simulações com o custo por unidade de cada estrato, considerando-se ainda os acessos e as distâncias às comunidades. Tal procedimento se deu pelo fato de não se dispor de resultados de pesquisas anteriores ou de uma amostra piloto que nos auxiliasse na tomada dessas decisões. Após várias simulações e com base em informações do IBGE e da Prefeitura Municipal, chegou-se a um consenso e obteve-se a alocação para estes estratos, conforme apresentado na Tabela 4.1.

Sendo:

- N_h Tamanho da população no estrato h ;
- n_h Tamanho da população entrevistada no estrato h ;
- c_h Custo por unidade observada no estrato h ;
- σ_h Variabilidade dentro do estrato h .

Tabela 4.1: Alocação ótima de Neyman

Estrato	N_h	c_h	σ_h	Alocação
1	25.043	1,50	2,0	871
2	9.283	2,00	2,0	278
3	3.982	2,00	2,0	120
4	9.825	2,00	2,5	370
5	9563	2,00	2,5	361
Total	57.651			2.000

Para a realização dessa pesquisa, contou-se com a participação de 3 bolsistas e alguns colaboradores diretos e espontâneos, assim distribuídos: Prof^a. D^a. Selene Maria Coelho Loibel (responsável direta pelo projeto), Karoline de Moraes, Murilo da Silva Nascimento e Sophia Lanza de Andrade (bolsistas), Olívio Crispim de Medeiros (colaborador direto) e Dacio Depercia além de outros cinco participantes (colaboradores espontâneos).

4.2.2 Análise Estatística

Nessa pesquisa foram realizadas 2.000 entrevistas de um total de 57.651 domicílios. O número de cães informados por ocasião dessa pesquisa foi de 1.810 animais, distribuídos em 1.131 domicílios, enquanto que em 869 destes, os informantes disseram não possuírem nenhum animal dessa espécie. Estima-se, através dessa pesquisa, o valor médio de 3,15 habitantes por domicílios, onde nessa área, com 2.000 domicílios entrevistados ter-se-á uma população estimada da ordem de 6.300 habitantes. Enquanto a OMS recomenda uma relação máxima de 1 cão para 7 humanos e a Fundação Nacional da Saúde recomenda a relação de 1 : 8, na cidade de Rio Claro, estima-se o número de 1 cão para cada 3,5 humanos, ou seja, uma situação com praticamente o dobro do recomendado por aquelas instituições.

O número total de domicílios por estrato e visitados encontram-se dispostos conforme Tabela 4.2.

Tabela 4.2: Distribuição do N° de domicílios por estratos e entrevistados

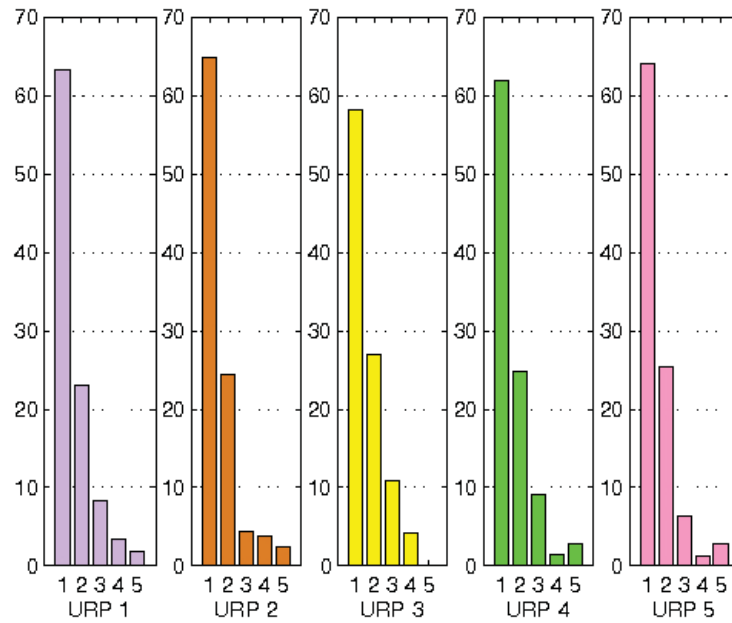
Estrato	N° total de Domicílios	N° de domicílios Entrevistados	Percentual de Cobertura
1	25.043	855	3,41
2	9.238	286	3,10
3	3.982	123	3,09
4	9.825	370	3,77
5	9.563	366	3,83
Total	57.651	2.000	3,47

O número total de domicílios entrevistados com o número de cães que possuem, encontram-se distribuídos na Tabela 4.3. Entre os domicílios que apresentam cães, essa informação poderá ser visualizada na Figura 4.1.

Tabela 4.3: Percentual do n° de domicílios com o n° de cães que possuem

Estrato	N° de cães que possuem - %						Total
	0	1	2	3	4	≥ 5	
1	20,1	14,3	5,2	1,9	0,8	0,4	42,7
2	6,3	5,2	2,0	0,4	0,3	0,2	14,4
3	2,4	2,1	1,0	0,4	0,1		6,0
4	8,0	6,5	2,6	1,0	0,1	0,3	18,5
5	6,5	7,6	3,0	0,8	0,1	0,4	18,4
Total	43,3	35,7	13,8	4,5	1,4	1,3	100,0

Figura 4.1: % do n° de domicílios com 1, 2, 3, 4, 5 ou mais cães

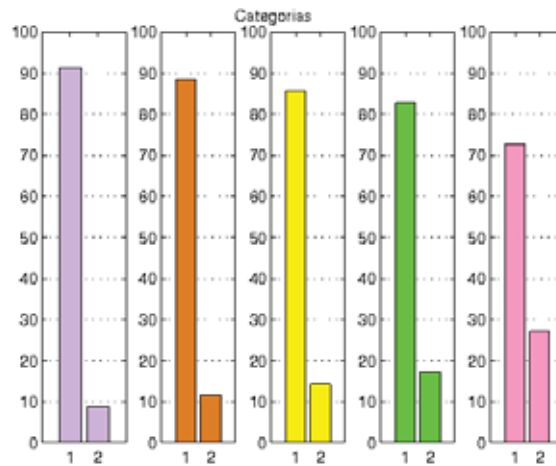


Do total de 1810 animais que foram cadastrados por ocasião das entrevistas, em 03 domicílios não foi respondida a questão sobre a categoria da OMS em que o cão se enquadra, conforme listado na Tabela 4.4, totalizando 7 cães.

Tabela 4.4: Porcentagem dos cães segundo a categoria da OMS a que pertencem

Estrato	Categoria OMS		Não Respondeu	Total
	1	2		
1	36,7	3,4	0,1	40,2
2	12,3	1,5		13,8
3	5,6	0,9		6,5
4	15,4	3,2		18,6
5	15,2	5,4	0,3	20,9
Total	85,2	14,4	0,4	100,0

Figura 4.2: Distribuição dos cães por categoria OMS - %



O total dos animais informados nessa pesquisa e, identificados os estratos a que pertencem, totalizando 1803 animais, encontram-se distribuídos na Tabela 4.5 segundo o sexo e o estrato a que pertencem. Esse número será a referência na construção de todas as tabelas que seguem.

Tabela 4.5: Total de cães distribuídos por estrato e sexo

Estrato	Machos - %	Fêmeas - %	Total
1	17,8	22,3	40,2
2	5,4	8,5	13,8
3	2,7	3,9	6,6
4	7,7	10,9	18,5
5	10,6	10,3	20,9
Total	44,1	55,9	100,0

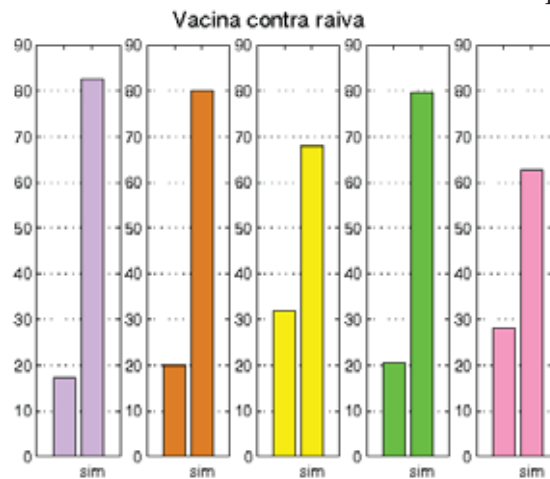
Os cães classificados nas categorias 1 e 2 segundo a OMS, encontram-se distribuídos na Tabela 4.6, separados pelo sexo e pelo estrato a que pertencem, da seguinte forma:

Tabela 4.6: Total de cães distribuídos por estrato, categoria e sexo

Estrato	Categoria 1 - %		Categoria 2 - %		Total
	Machos	Fêmeas	Machos	Fêmeas	
1	15,7	21,1	2,2	1,2	40,2
2	4,6	7,7	0,7	0,8	13,8
3	2,0	3,4	0,7	0,5	6,6
4	5,8	9,7	1,9	1,3	18,7
5	7,3	8,0	3,2	2,2	20,7
Total	35,4	49,9	8,7	6,0	100,0

O percentual de animais vacinados, por estratos, é mostrado na Figura 4.3.

Figura 4.3: % de animais vacinados contra raiva por estrato



O controle de natalidade entre os cães, referente a castração, por estrato, é apresentado na Figura 4.4.

Para as fêmeas, além da castração, existem ainda outros métodos para o controle de natalidade, conforme apresentados na Figura 4.5. Assim, em cada estrato, tem-se os percentuais dos entrevistados que disseram tomar as seguintes providências com as fêmeas no período do cio: a coluna 1, representa as pessoas que responderam não fazerem nada; a coluna 2, fêmeas mantidas presas por ocasião do cio; a coluna 3, informaram que a sua fêmea era filhote e tomariam as devidas providências na época oportuna; e finalmente, a coluna 4, representando os entrevistados que responderam fazerem aplicações de injeções.

Figura 4.4: Controle de Natalidade - %

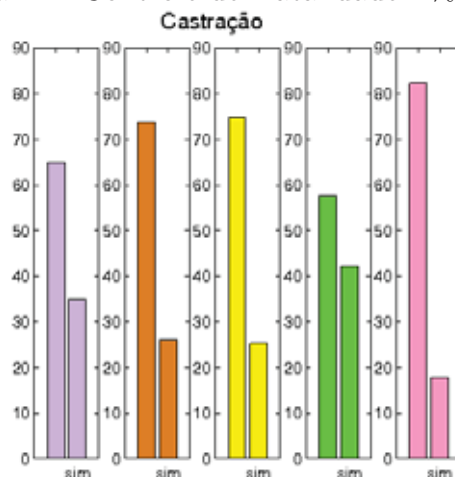
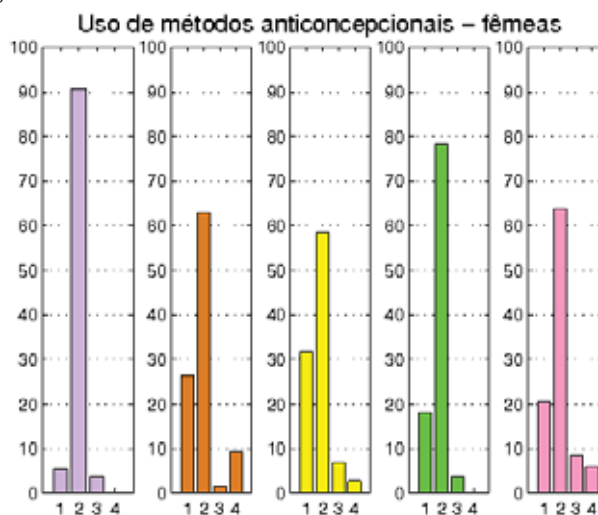


Figura 4.5: Controle de Natalidade das Fêmeas - %



4.2.2.1 Análise dos cães da Categoria 1

Os cães da categoria 1, que só saem as ruas quando acompanhados de seus respectivos donos, que aparentemente não causam maiores problemas à saúde pública, apresentam os seguintes resultados (ver Tabela 4.7), no item que se refere a "Vacinação contra Raiva".

No quesito Vacinação contra Raiva, entre os animais machos, a situação mais crítica está no estrato 1 com um percentual de 2,8% de animais não vacinados, em particular na comunidade Jardim Araucária, com um percentual de 36,4%. Também no estrato 1, está a maior concentração de animais não vacinados entre as fêmeas, com um percentual de 4,4% de animais desprotegidos, sendo o Alto do Santana, Jardim do Trevo e a Vila Operária os maiores responsáveis pelos animais não vacinados, representando

Tabela 4.7: Cães da categoria 1, Vacinação, por estrato e sexo

Estrato	Machos - %			Fêmeas - %			Total
	Sim	Não	Não Resp.	Sim	Não	Não Resp.	
1	15,5	2,8		20,3	4,4		43,0
2	4,1	1,3		7,1	1,8	0,1	14,4
3	1,7	1,0		3,3	0,6		6,6
4	5,8	1,0		9,5	1,7	0,1	18,1
5	3,7	1,3	3,6	3,8	2,3	3,2	17,9
Total	30,8	7,4	3,6	44,0	10,8	3,4	100,0

percentuais de 24,0%, 24,0% e 21,6% entre os animais de cada comunidade respectivamente.

Ainda no estrato 1, no quesito Vacinação, com 100% de cobertura vacinal, merecem destaque as comunidades Cidade Jardim, Jd. Claret, Jd. Boa Vista, Jd. Mirassol, Jd. Santa Eliza, Pq. Universitário e Zona Central, entre os machos, e as comunidades Cidade Jardim, Jd. Itapuã, Jd. Mirassol e Pq. São Paulo, entre as fêmeas.

Para os cães da categoria 1, no item que se refere à castração, para os machos tem-se os seguintes resultados:

Tabela 4.8: Cães da categoria 1, Machos por estrato, Castração

Estrato	Animais Castrados - %		Total
	Sim	Não	
1	11,3	32,6	43,9
2	2,8	10,2	13,0
3	1,2	5,1	6,3
4	7,5	8,7	16,2
5	3,0	17,6	20,6
Total	25,8	74,2	100,0

Para as fêmeas, o controle de natalidade na categoria 1 se dá segundo os critérios apresentados na Tabela 4.9, sendo eles: castração, fêmeas presas por ocasião do cio, aplicação de injeção, filhotes e sem controle de natalidade.

No item Controle de Natalidade, a situação mais crítica entre os machos, também se encontra no estrato 1, com 32,6% de animais não castrados, sendo o bairro Santana o maior responsável por esse índice, que atinge o percentual de 93,5% entre os animais de

Tabela 4.9: Cães da categoria 1, Fêmeas por estrato, Controle de Natalidade

Estrato	Controle de Natalidade - %						Total
	Cast.	Presa	Inj.	Filhote	S/ Cont.	N. Resp.	
1	18,1	22,6	0	1,3	0,4		42,4
2	4,1	7,0	1,7	0,2	2,1	0,2	15,3
3	1,7	2,8	0,2	0,6	1,6		6,9
4	9,1	8,3	0	0,6	1,5		19,5
5	2,5	10,1	0,8	1,1	1,3	0,1	15,9
Total	35,5	50,8	2,7	3,8	6,9	0,3	100,0

sua comunidade. Com relação as fêmeas, a situação mais crítica acontece no estrato 2, onde constata-se o percentual de 2,1% de animais sem nenhum controle de natalidade, sendo o Bairro Cherveson o maior responsável por esse índice com 35,3% dos animais de sua comunidade.

4.2.2.2 Análise dos Cães da Categoria 2

Os cães da categoria 2, que podem sair às ruas a qualquer dia e a qualquer horário, sem um acompanhamento efetivo de seus proprietários, que a princípio são os cães que causam maiores problemas à saúde pública, juntamente com os cães das categorias 3 e 4, aqui não mencionados e ralatados, apresentam os seguintes resultados (Tabela 4.10, no item que se refere a "Vacinação contra Raiva").

Tabela 4.10: Cães da categoria 2, critério de Vacinação, por estrato e sexo

Estrato	Machos - %			Fêmeas - %			Total
	Sim	Não	Não Resp.	Sim	Não	Não Resp.	
1	11,1	3,8		5,0	3,4		23,3
2	4,6	0,4		4,6	1,1		10,7
3	1,5	1,5		2,3	1,1		6,4
4	6,5	6,9		6,5	2,3		22,2
5	8,8	6,1	6,9	4,6	1,9	9,1	37,4
Total	32,5	18,7	6,9	23,0	9,8	9,1	100,0

No item vacinação, entre os animais machos, a pior situação ocorre no estrato 4, onde constata-se o Jardim Flórida como sendo o maior responsável por esse índice

com 78,6% de seus animais não vacinados. Já entre as fêmeas, a situação mais crítica continua sendo no estrato 1, onde destaca-se as comunidades Jardim Mirassol, Jardim Olinda e o Bairro Santana que atingem 100% de seus animais não vacinados.

Para os cães dessa categoria, que é favorecido o livre acasalamento entre os animais dessa espécie, provocando assim um aumento cada vez maior na população, no item castração dos machos tem-se os resultados, conforme a Tabela 4.11

Tabela 4.11: Cães da categoria 2, Machos por estrato, Castração

Estrato	Animais Castrados - %		Total
	Sim	Não	
1	5,9	19,7	25,6
2	2,6	5,9	8,5
3	0,7	4,6	5,3
4	3,3	19,7	23,0
5	4,0	33,6	37,6
Total	16,5	83,5	100,0

Ainda nessa categoria, para as fêmeas, o controle de natalidade se dá segundo os métodos especificados na Tabela 4.12, que são os mesmos descritos para a Tabela 4.9.

Tabela 4.12: Cães da categoria 2, Fêmeas por estrato, Controle de natalidade

Estrato	Controle de Natalidade - %					Total
	Cast.	Presa	Inj.	Filhote	S/ Cont.	
1	9,1	0	0	0	10,9	20,0
2	5,5	0	2,7	0	5,5	13,7
3	4,6	0	0	0	3,6	8,2
4	6,4	1,8	0	0	12,7	20,9
5	11,8	9,1	2,7	0,9	12,7	37,2
Total	37,4	10,9	5,4	0,9	45,4	100,0

No tocante ao controle de natalidade, entre os machos e no quesito castração, a pior situação acontece no estrato 5, sendo o Jd. Inocoop o maior responsável por esse índice, com 66,7% de seus animais não castrados. Entre as fêmeas as piores situações ocorrem nos estratos 4 e 5 com os percentuais de 12,7% em cada estrato de animais sem nenhum controle de natalidade, sendo os bairros Cidade Nova, Indaiá, Jardim América, Jardim dos Ipês e Vila Nova no estrato 4 e os bairros Jardim Novo e Novo Wenzel no estrato 5, onde nestas comunidades 100% destes animais são totalmente desprotegidos nesse quesito.

4.2.2.3 Cães de Vizinhaça

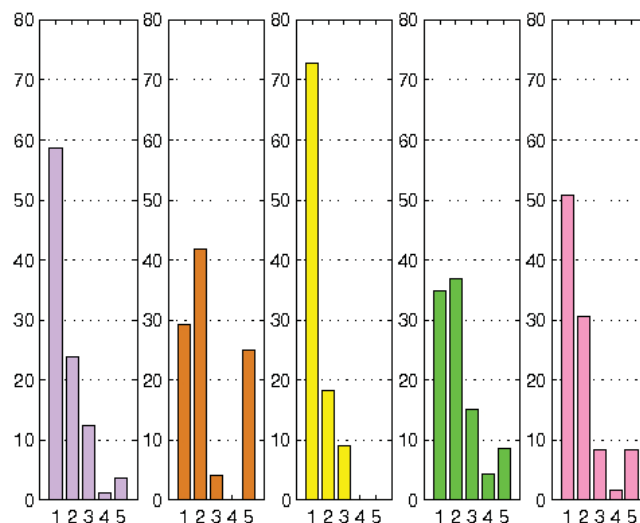
Para os cães de vizinhaça que irão formar a categoria 3, a informação que se dispõe, tendo em vista a impossibilidade de sua contagem, é que de acordo com as informações obtidas no inquérito domiciliar existem 347 animais recebendo assistência, sendo que 31,4% recebem assistência de pessoas que não possuem cães e 68,6% recebem de pessoas que já possuem esses animais.

Entre as pessoas que dão assistência a cães de vizinhaça tem-se a distribuição conforme Tabela 4.13 ou a Figura 4.6.

Tabela 4.13: Cães de vizinhaça que recebem assistência

Nº de Cães Beneficiados	1	2	3	4	5 ou mais
% de Domicílios	49,0	30,0	10,9	1,8	8,2

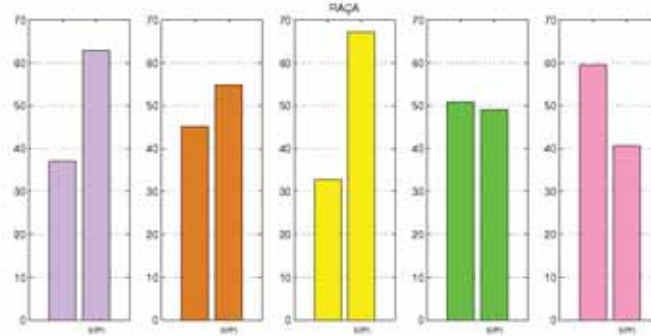
Figura 4.6: % de domicílios cujas pessoas dão assistência a 1, 2, 3, 4, 5 ou mais cães



A característica dos cães no tocante à raça, por estratos, é apresentado na Figura 4.7.

Outros detalhes como, Porte, Faixa Etária, etc., estão apresentados nas tabelas do Anexo B "Dados da População Canina".

Figura 4.7: Percentual dos cães de raça



4.2.2.4 Comparação AAS_S x AE

Nesse item será comparado os resultados da pesquisa com os métodos de amostragem AAS_S e AE . Parte dos cálculos apresentados na Tabela 4.14 foram efetuados pelo programa Statistica e parte destes utilizando-se o microsoft Excel e o Matlab. Para a sua análise tomou-se \bar{y}_h : média do nº de cães por domicílio, por estrato, o intervalo de confiança para a média e também a variância e o desvio padrão na amostra, além dos coeficientes de variação por estratos.

Tabela 4.14: Resumo dos cálculos por estratos

Est.	Nº Dom.	Nº Cães	\bar{y}_h	dp	$\widehat{V}(\bar{y}_h)$	IC p/μ	CV
1	855	727	0,8503	1,1333	0,0015	(0,77; 0,93)	1,33
2	286	250	0,8741	1,1168	0,0044	(0,74; 1,00)	1,28
3	123	119	0,9675	1,0238	0,0085	(0,78; 1,15)	1,06
4	370	336	0,9081	1,0856	0,0032	(0,80; 1,02)	1,20
5	366	378	1,0328	1,2688	0,0044	(0,90; 1,16)	1,23
Total	2.000	1.810	0,9050	1,1431	0,0007	(0,85; 0,96)	-

O cálculo para amostra, com os critérios da AAS_S , pode ser visualizado na linha referente ao item "Total". Na sequência calcula-se alguns resultados para a amostra com critérios da AAS_S . A seguir algumas estimativas por estrato e para a amostra toda para a comparação dos resultados. Como $\frac{n}{N} = \frac{2000}{57.651} = 0,035$, considera-se que a população é infinita.

Assim, para a ASS_G tem-se:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n (Y_i) = 0,9050$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{y})^2 = 1,3066$$

$$\widehat{V}(\bar{y}) = \frac{s^2}{n} = 0,0007$$

Considerando um coeficiente de confiança $1 - \alpha = 0,95$, tem-se da tabela da distribuição normal que o valor de $z_{\alpha/2} = 1,96$. Assim:

$$IC(\mu) = \bar{y} \pm z_{(\alpha/2)} \sqrt{\widehat{V}(\bar{y})} = 0,9050 \pm 0,0519 = (0,85 ; 0,96)$$

A estimativa do total populacional é dada por

$$\hat{\tau} = \frac{N}{n} \sum_{i=1}^n Y_i = 52.174,$$

cuja variância é estimada por

$$\widehat{V}(\hat{\tau}) = N^2 \frac{s^2}{n} = 2.326.546.$$

O intervalo de confiança para τ é estimado por

$$IC(\tau) = \hat{\tau} \pm z_{(\alpha/2)} \sqrt{N^2 \frac{s^2}{n}} = 52.174 \pm 2.989 = (49.185; 55.164).$$

Para a amostragem estratificada, apresenta-se os cálculos apenas para o estrato 1, sendo os demais cálculos efetuados de forma semelhante. Em seguida, calcula-se alguns resultados para toda a amostra, condensada pelos estratos.

$$\bar{y}_1 = \frac{1}{n_1} \sum_{n=1}^{n_1} Y_{1i} = 0,8503$$

$$s_1^2 = \frac{1}{n_1-1} \sum_{n=1}^{n_1} (Y_{1i} - \bar{y}_1)^2 = 1,2843$$

$$\widehat{V}(\bar{y}_1) = \frac{s_1^2}{n_1} = 0,0015$$

$$IC(\mu_1) = \bar{y}_1 \pm z_{(\alpha/2)} \sqrt{\widehat{V}(\bar{y}_1)} = 0,8503 \pm 0,0759 = (0,77; 0,93).$$

Para a amostra condensada, tem-se os cálculos:

$$\bar{y}_{es} = \frac{1}{N} \sum_{h=1}^H N_h \hat{\mu}_h = 0,9023$$

$$\widehat{V}(\bar{y}_{es}) = \sum_{h=1}^H W_h^2 V(\hat{\mu}_h) = 0,0006$$

$$IC(\mu) = \bar{y}_{es} \pm z_{\alpha/2} \sqrt{\widehat{V}(\bar{y}_{es})} = 0,9023 \pm 0,0496 = (0,85; 0,95)$$

$$\hat{\tau}_{es} = \sum_{h=1}^H \sum_{n=1}^{n_h} Y_{hi} = 52.020$$

$$\widehat{V}(\hat{\tau}_{es}) = \sum_{h=1}^H N_h^2 \frac{s_h^2}{n_h} = 2.162.288$$

$$IC(\tau_{es}) = \tau_{es} \pm z_{(\alpha/2)} \sqrt{\widehat{V}(\hat{\tau}_{es})} = 52.020 \pm 2.882 = (49.138; 54.902).$$

Conforme pode ser notado pelo cálculos apresentados, principalmente pelos intervalos de confiança das médias e dos totais, não há diferença significativa quando da utilização de uma ou outra técnica de amostragem. Tais fatos podem ser justificados pelo mau dimensionamento dos estratos, que segundo dados da Prefeitura e do IBGE, os mesmos foram estabelecidos de acordo com o nível sócio-econômico da população, mas de fato, as URP's foram determinadas por outros critérios, como pode ser visto durante a pesquisa. Pela falta de informações sobre alguma variável que poderia ser usada para a estratificação, optou-se por utilizar a divisão em URP.

4.2.3 Estimação das Proporções

Para o caso específico deste projeto, onde só foi possível identificar duas das categorias, com 1.541 animais na categoria 1 e 262 na categoria 2, num total de 1.803 cães, tem-se para estes casos as seguintes proporções amostrais:

$$\hat{P}_1 = \frac{\sum Y_i}{n} = 0,85.$$

$$\hat{P}_2 = 1 - \hat{P}_1 = 0,15.$$

Uma estimativa da variância para \hat{P} é dada por

$$\widehat{V}(\hat{P}) = \frac{\hat{P}(1 - \hat{P})}{n} = 0,00007.$$

Os intervalos de confiança, com coeficiente de confiança $1 - \alpha$ para \hat{P}_1 e \hat{P}_2 são dados por

$$IC(\hat{P}_1, 1 - \alpha) = \hat{P}_1 \pm z_{(\alpha/2)} \sqrt{\widehat{V}(\hat{P})} = (0,83; 0,87)$$

e

$$IC(\hat{P}_2, 1 - \alpha) = \hat{P}_2 \pm z_{(\alpha/2)} \sqrt{\widehat{V}(\hat{P})} = (0,13; 0,17).$$

5 Conclusão

De fato, a amostragem estratificada produz resultados mais eficientes do que a amostragem aleatória simples, desde que realmente ao se usar a primeira técnica, os estratos sejam muito bem dimensionados segundo a variável escolhida, conforme demonstra o exemplo de aplicação mencionado no final do capítulo 3. Pode-se então propor esse plano como sendo realmente o mais preciso, com os melhores estimadores e indicado principalmente para as populações assimétricas, que é o caso da população canina de Rio Claro, em estudo nesse trabalho.

Mais uma vez ressalta-se que, para os estratos usados no projeto de extensão "Dimensionamento da População Canina de Rio Claro - SP", com informações fornecidas pelo IBGE e Prefeitura Municipal de Rio Claro, que segundo esses órgãos esses estratos foram formados segundo os critérios de renda entre os seus habitantes, esse fato não condiz com a realidade como pode ser visto nos resultados apresentados.

Vale ressaltar ainda, a necessidade da continuidade deste projeto de extensão, para que com toda a experiência adquirida no decorrer do ano de 2012, por ocasião das coletas dos dados em campo, se possa melhor estratificar a população, tomando os dados do primeiro ano como uma amostra piloto. Pôde-se verificar de perto, o diferencial que existe no nível social da população humana associada com a presença de cães em suas residências. Fica também aqui registrado o fato da amostragem estratificada apresentar custos mais elevados do que a amostragem aleatória simples, pois para a aplicação da primeira técnica, há que se dimensionar todos esses estratos tendo como base o poder aquisitivo das pessoas, que a princípio, parece ser a melhor variável para a estratificação.

Como um fruto muito valioso deste presente trabalho, destaca-se a oportunidade de manipulação de alguns programas que apresentam a linguagem \TeX .

Destaca-se ainda os vários conhecimentos adquiridos no campo da amostragem estatística, além da vivência com a prática de coleta de dados para o projeto de extensão acima epigrafado.

Referências

- [1] ALVES, M. C. G. P. et al. Dimensionamento da população de cães e gatos do interior do estado de são paulo. *Revista Saúde Pública*, v. 39(6), p. 891 – 897, 2005.
- [2] BOLFARINE, H.; BUSSAB, W. de O. *Elementos de Amostragem*. 1. ed. São Paulo: Edgard Blucher, 2005.
- [3] BRITO, J. et al. An exact algorithm for the stratification problem with proportional allocation. *Optimization Letters*, v. 4 n° 2, p. 185 – 195, 2010.
- [4] BUSSAB, W. de O.; MORETIN, P. A. *Estatística Básica*. 8. ed. São Paulo: Saraiva, 2013.
- [5] COCHRAN, W. G. *Técnicas de Amostragem*. 1. ed. Rio de Janeiro: Editora Fundo da Cultura, 1965.
- [6] CRUZ, J. H. da. *Amostragem Estatística: Noções Básicas*. 1. ed. Aracaju, SE: Universidade Federal de Sergipe, 1978.
- [7] DALENIUS, T.; HODGES, J. L. J. Minimum variance stratification. *Journal of the American Statistical Association*, v. 54 n° 285, p. 88 – 101, 1959.
- [8] HORGAN, J. M. Choosing the stratification boundaries: The elusive optima. *Istanbul University Journal of the school of Business Administration*, v. 39 n° 2, p. 195 – 204, 2010.
- [9] GUNNING, P.; HORGAN, J. M.; KEOGH, G. Obtaining stratum breaks in skewed populations using a simple method. *ASA Section on Survey Research Methods*, p. 3620 – 3626, 2004.
- [10] MAGALHÃES, M. N.; LIMA, A. C. P. de. *Noções de Probabilidade e Estatística*. 7. ed. São Paulo: Editora da Universidade de São Paulo, 2011.
- [11] MATOS, M. R. de et al. Técnica pasteur são paulo para dimensionamento de população canina. *Caderno Saúde Pública*, v. 18(5), p. 1423 – 1428, 2002.
- [12] MEMÓRIA, J. M. P. *Breve História da Estatística*. 1. ed. Brasília: Embrapa Informação Tecnológica, 2004.

- [13] MUNIZ, J. A.; ABREU, A. R. de. *Técnicas de Amostragem (Coleção Textos Acadêmicos)*. 1. ed. Lavras, MG: Ufla/Faepe, 1999.
- [14] SOTO, F. R. M. et al. Dinâmica populacional canina no município de Ibiúna - SP: Estudo retrospectivo. *Braz. J. vet. anim. Sci*, v. 43 n° 2, p. 178 – 185, 2006.
- [15] UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE. História da Estatística. Natal, RN: Disponível em: <<http://www.estadistica.ccet.ufrn.br/historia.php>>. Acesso em: 21/03/12.

A Dados de Áreas cultivadas com Milho em Lavras - MG

A.1 Dados para estimação de uma amostra aleatória simples e estratificada

Tabela A.1: Área cultivada com milho, em ha, de 194 propriedades de Lavras - MG, na safra de 1988

40,2	4,5	5,1	1,4	1,2	1,5	12,0	22,0	125,4	8,3	25,0	53,0	9,2
7,5	1,9	1,0	3,7	20,8	62,0	7,0	28,2	72,0	5,3	8,9	1,2	1,0
1,6	11,4	34,0	97,0	13,0	28,0	54,0	5,8	8,1	1,2	1,2	1,8	14,0
28,3	1,8	12,4	27,0	64,0	7,3	6,3	1,5	2,4	3,6	17,3	1,3	1,2
16,6	28,6	82,0	7,3	4,2	1,4	0,8	1,3	1,2	2,4	3,9	12,0	31,0
88,0	5,7	7,5	1,3	1,6	8,2	1,7	1,3	1,4	18,2	38,4	85,0	4,8
6,3	1,3	6,4	8,2	1,2	1,5	1,0	12,8	30,0	88,0	5,4	7,2	62,0
7,6	4,9	2,4	1,0	1,0	15,4	39,6	90,0	6,9	39,0	10,1	8,5	4,2
1,8	1,4	3,9	14,6	32,0	75,0	7,8	13,0	34,0	8,2	8,8	15,7	37,3
4,5	7,4	15,4	33,0	19,3	29,7	14,8	38,0	14,2	37,0	17,8	29,0	17,3
6,1	1,2	1,8	1,2	11,0	1,8	15,0	38,0	87,0	6,8	4,8	1,4	1,2
1,6	10,3	3,8	1,5	12,2	35,0	112,0	6,5	9,0	1,8	3,6	3,8	1,3
1,2	1,5	17,0	32,0	74,0	4,7	4,9	0,7	1,1	5,9	1,6	1,8	1,8
15,7	35,3	83,0	9,7	5,4	1,2	7,3	7,4	15,0	34,0	5,0	11,4	27,0
96,0	5,7	7,5	34,9	7,3	16,2	35,0	14,3	19,8	13,8	13,7	15,6	-

Fonte: Muniz, 1999.

B Dados Gerais da População Canina

Questionário utilizado nas entrevistas dos domicílios

ESTRATO: _____ PESQUISADOR: _____

QUESTIONÁRIO N°: _____

1. O(A) Sr(a). possui cães em () casa / () apto? () Sim. Quantos? ____ () Não
(Se não, vá direto p/ a questão 4)

2. Preencher a tabela abaixo referente a cada cão da casa (apto)

Cães	1	2	3	4	5	6	7	...	17
Sexo								...	
Idade								...	
Porte								...	
Raça								...	
Vacina (raiva)								...	
Categoria								...	

(SE SO TEM MACHOS, PULAR PARA QUESTÃO 4)

3. Você utiliza algum outro método para evitar que suas cadelas tenham filhotes?

4. O(A) Sr(a). alimenta e/ou dá alguma assistência a cães da vizinhança? () Sim.
Quantos? _____ () Não

5. O(A) Sr(a). sabe se mais alguém da rua cuida desses cães?

() Sim. Quem? _____ () Ninguém () Não sei

Obs.: Explicar sobre as categorias 1- cães que só saem com dono, 2- saem sozinhos ou 3- cães de vizinhança (entregar a coleira da cor referente à categoria para que não sejam contado 2 vezes)

Dados da População Canina

Tabela B.1: Percentuais da População Canina - Machos

Parâmetros	Especificações	Estrato 1	Estrato 2	Estrato 3	Estrato 4	Estrato 5
Castrados	Sim	12,6	2,8	1,1	9,0	3,1
	Não	27,8	9,4	5,0	8,4	20,8
Vacina c/ Raiva	Sim	33,8	9,5	3,8	13,1	11,3
	Não	6,6	2,6	2,4	4,3	3,3
	Não Resp.					9,4
Raça	Sim	26,5	5,9	4,5	9,3	9,5
	Não	13,9	6,3	1,6	8,1	14,3
	Não Resp.					0,1
Porte	1	18,3	4,9	2,4	7,0	7,8
	2	13,9	5,8	2,1	5,9	10,3
	3	8,3	1,5	1,6	4,5	5,9
Categoria OMS	1	35,4	10,5	4,5	13,0	16,5
	2	4,9	1,6	1,6	4,4	7,1
	Não Resp.	0,1				0,3
Faixa Etária (Em anos)	≤ 1	5,0	2,6	0,9	3,1	5,8
	1 a 3	12,6	4,0	1,9	4,1	7,0
	3 a 5	8,3	1,9	1,6	3,0	4,4
	5 a 7	4,4	1,0	0,5	2,0	2,3
	7 a 9	3,3	0,8	0,3	1,8	1,6
	9 a 11	2,8	0,5	0,5	1,8	1,3
	11 a 13	1,9	0,6	0,1	0,8	0,6
	13 a 15	0,9	0,1	0,3	0,4	0,3
	> 15	0,9			0,1	
Não Resp.	0,5	0,6	0,1	0,4	0,8	

Tabela B.2: Percentuais da População Canina - Fêmeas

Parâmetros	Especificações	Estrato 1	Estrato 2	Estrato 3	Estrato 4	Estrato 5
Controle de Natalidade	Castradas	17,4	4,3	2,0	8,8	4,2
	Presas	20,0	6,2	2,5	7,5	9,3
	Injeção		2,0	0,2	0,2	1,0
	Filhotes	1,3	0,2	0,5	0,3	1,2
	S. Controle	1,0	2,3	1,8	2,7	2,6
	Não Resp.	0,3	0,2			0,3
Vacina c/ Raiva	Sim	32,3	12,1	5,0	15,9	7,2
	Não	7,6	3,0	1,9	3,5	3,0
	Não Resp.		0,1		0,1	8,3
Raça	Sim	24,2	8,9	4,4	9,0	7,4
	Não	15,7	6,2	2,6	10,5	11,1
Porte	1	18,2	6,3	3,3	9,4	7,2
	2	14,7	6,8	2,6	7,7	8,3
	3	7,0	2,0	1,0	2,4	3,0
	Não Resp.			0,1		
Categoria OMS	1	37,7	13,6	6,0	17,2	14,3
	2	2,2	1,5	0,9	2,3	3,9
	Não Resp.	0,1				0,3
Faixa Etária (Em anos)	≤ 1	4,6	2,7	1,0	2,5	3,7
	1 a 3	10,5	3,4	2,4	5,9	4,9
	3 a 5	7,0	3,3	1,5	3,0	3,4
	5 a 7	5,9	2,3	0,4	3,4	2,0
	7 a 9	4,6	1,4	0,4	2,1	0,9
	9 a 11	3,3	0,7	0,9	0,8	1,7
	11 a 13	1,6	0,5	0,1	1,3	0,5
	13 a 15	1,3	0,5	0,1	0,4	0,2
	> 15	0,5	0,2		0,1	0,2
Não Resp.	0,6	0,3	0,2	0,1	1,1	