

Plano amostral e ponderação pelo efeito de desenho de um levantamento epidemiológico de saúde bucal

Sample plan and weighting by design effect of the epidemiological survey of oral health

Andréa Maria Eleutério Barros de Lima Martins¹, Pedro Eleutério dos Santos-Neto², Leandro Henrique Souza Batista³, Jairo Evangelista Nascimento⁴, Allyson Ferreira Gusmão⁵, Núbia Barbosa Eleutério⁶, André Luiz Sena Guimarães⁷, Alfredo Maurício Batista de Paula⁸, Desirée Sant'Ana Haikal⁹, Marise Fagundes Silveira¹⁰, Isabela Almeida Pordeus¹¹

Resumo: **Objetivo:** Descrever o planejamento amostral, métodos de estimação e ponderação. **Metodologia:** Garantiu-se representatividade de 6 estratos etários, buscando o maior tamanho de amostra ao se considerar proporções (50%), médias de dentes cariados, perdidos e obturados (CPOD), desvio padrão (DP) prévio, *deff* igual a 2,0 e taxa de não resposta de 20%. Optou-se por uma amostra complexa probabilística por conglomerados e propõe-se ponderação pelo efeito de desenho. **Resultados:** A amostra planejada, considerando proporções, foi de 4478 e de 7551, considerando o CPOD/DP, sendo a final de 4852. Dentre os 276 setores censitários urbanos, sortearam-se 52 e dentre as 11 áreas rurais, sortearam-se duas. No segundo estágio, em cada um dos 52 setores, sorteou-se de uma a 11 quadras, totalizando 354 quadras. A taxa de não resposta foi próxima a 10% nos estratos, avaliando-se 4509 pessoas. O peso para cada conglomerado variou de 3,7 a 47,1. **Conclusões:** Após ponderação, constataram-se modificações nas proporções da amostra. A amostragem e a ponderação apresentadas foram baseadas em propostas recentes, garantindo poder de inferência necessário às investigações epidemiológicas.

Palavras-chave: Amostragem por conglomerados. Estudos Populacionais em Saúde Pública. Técnicas de estimação. Análise estatística. Saúde bucal.

Abstract: **Objective:** To describe sample design, weighting and estimation methods. **Methodology:** Representation of six age groups was ensured seeking the largest sample size when considering proportions (50%), average number of teeth decayed, missing and filled (DMF), standard deviation (SD) prior, *deff* equals to 2.0 and non-response rate of 20%. A random probabilistic sample by conglomerates was chosen and weighting for the design effect proposed. **Results:** The planned sample, considering the proportions, was of 4478 and of 755, considering the DMFT/SD, being the final one 4852. From the 276 urban census tracts 52 were raffled and from among the 11 rural areas two were raffled. At the second stage, in each of the 52 sectors 1 to 11 blocks were raffled, totaling 354 blocks. The non-response rate was close to 10% in the strata, assessing 4509 people. The weight obtained for each conglomerate ranged from 3.7 to 47.1. **Conclusion:** After weighting, changes were found in the proportions of the sample. Sampling and weighting presented were based on recent proposals, ensuring necessary power of inference to epidemiological investigations.

Key-words: Cluster Sampling. Population Studies in Public Health. Estimation Techniques. Statistical Analysis. Oral Health.

-
- 1 Doutora em Saúde Pública / Epidemiologia - Universidade Federal de Minas Gerais - UFMG. Professora da Unimontes.
 - 2, 4 Mestre em Ciências da Saúde - Unimontes. Professor da Unimontes.
 - 3, 5 Bacharel em Matemática - Unimontes.
 - 6 Pós-graduada em Residência Multiprofissional em Saúde da Família - Unimontes.
 - 7 Pós-doutor em Biologia Celular - University of Western Ontario. Professor da Unimontes.
 - 8 Doutor em Patologia - UFMG. Professor da Unimontes.
 - 9 Mestre em Odontologia em Saúde Coletiva - UFMG. Professora da Unimontes.
 - 10 Doutoranda no Programa de Pós-graduação em Saúde Coletiva - Universidade Federal de São Paulo - UNIFESP.
 - 11 Doutora em Epidemiology and Public Health - University College London. Professora titular em Odontopediatria da UFMG.

INTRODUÇÃO

Pesquisas epidemiológicas abrangem um amplo universo de indivíduos tornando-se, muitas vezes, impossível ou inviável considerá-los em sua totalidade. Por essa razão, é frequente trabalhar com uma amostra, ou seja, com uma parte representativa do universo.¹ Amostragem é, então, o processo pelo qual se extrai uma ou mais amostras de uma população de interesse e as observa com a finalidade de se estimar parâmetros para a população. O plano amostral é o conjunto de técnicas utilizadas para cálculo do tamanho amostral e seleção da amostra.^{1,2} Uma boa amostragem ocasiona custo reduzido, maior rapidez, maior amplitude e exatidão na investigação, fatores essenciais para o sucesso da pesquisa.^{3,4} Para se estabelecer um plano amostral, os objetivos da pesquisa e a população a ser amostrada devem ser bem definidos, bem como os parâmetros que precisam ser estimados para se atingir tais objetivos. Além disso, deve-se definir a unidade de amostragem e a forma de seleção dos elementos da população.⁵

O uso da amostragem probabilística, em que cada elemento da população possui uma probabilidade conhecida e diferente de zero de ser selecionado, é frequente na literatura.^{5,6} Os métodos probabilísticos vão desde o mais simples (amostragem aleatória simples - AAS) até os mais complexos, entre eles, a amostragem estratificada e a por conglomerados.⁷ Na AAS, todos os elementos da população apresentam a mesma probabilidade de serem amostrados. O fato de um elemento pertencer ou não à amostra torna-se, então, obra do acaso. Normalmente, a AAS é obtida por sorteio simples (fichas numeradas) ou através da utilização de programas de computador ou de tabelas de números aleatórios, comumente encontradas em livros de estatística.⁸ Já a amostragem estratificada consiste em se dividir a população em subgrupos homogêneos para determinadas características e selecionar uma amostra em cada um deles, separadamente. Nesse caso, cada uma das subdivisões populacionais é denominada de estrato. Esse tipo de amostragem é recomendado quando

se deseja obter estimativas com certa precisão para cada um dos estratos, diminuindo-se a possibilidade de erro.^{7,9}

Já nas amostras por conglomerados, a unidade amostral é um conjunto de elementos da população (conglomerado ou *cluster*). Este tipo de amostragem é repetida em múltiplos estágios, que vão sendo selecionados até se chegar ao objeto de interesse da pesquisa.⁹ Sua utilização é frequente nos estudos populacionais pela facilidade e diminuição no custo por elemento amostrado, já que se evitam gastos com elaboração de cadastros ou listas dos elementos da população, dispensando a localização de indivíduos específicos.^{5,7,9} Por outro lado, ao se investigar apenas os conglomerados sorteados, há aumento da imprecisão devido às possíveis correlações das unidades amostrais dentro e entre os conglomerados, fazendo-se necessário aumentar o tamanho da amostra. Além disso, há necessidade de se tratar, de forma apropriada, os dados para análises, pois os mesmos não podem ser tratados como se fossem observações independentes e identicamente distribuídas, como ocorre na amostragem aleatória simples.^{1,9}

Quanto à necessidade de aumento do tamanho da amostra, ao se utilizar conglomerados, torna-se indispensável a correção por um valor conhecido como *deff* (*design effect* = efeito de desenho), razão entre as imprecisões associadas à estimação usando uma AAS como referência. Utiliza-se a teoria da AAS e, posteriormente, corrige-se o valor encontrado multiplicando-o pelo valor do *deff*. Na prática, considera-se 1,4 ou 1,5 (correção no tamanho amostral entre 40% e 50%) para resguardar a precisão¹, quanto ao tratamento de forma apropriada dos dados para análise, devido às diferentes probabilidades de inclusão de cada conglomerado⁸, objeto de estudo e discussão nas literaturas nacional e internacional.^{7,9-11}

Tais medidas são necessárias, visto que algumas publicações apontam problemas que ocorrem quando a análise estatística não incorpora

MARTINS, A. M. E. B. L.; SANTOS-NETO, P. E.; BATISTA, L. H. S.; NASCIMENTO, J. E.; GUSMÃO, A. F.; ELEUTÉRIO, N. B.; GUIMARÃES, A. L. S. G.; PAULA, A. M. B.; HAIKAL, D. S.; SILVEIRA, M. F.; PORDEUS, I. A.

a estrutura do plano amostral.^{7,12} Para se compensar as probabilidades desiguais de seleção em cada estágio, nas amostras por conglomerados, devem-se atribuir ponderações diferenciadas aos elementos amostrais, que correspondem ao inverso dos produtos das probabilidades de inclusão nos diversos estágios de seleção.⁹ Ao se adotar a análise estatística tradicional, podem-se produzir incorreções que comprometem os resultados, os testes de hipóteses e as conclusões da pesquisa. Tal situação, por exemplo, ocorreu no Projeto SBBrazil 2002/2003¹³⁻¹⁵, de forma que tem sido sugerido que seus resultados não permitam inferências estatisticamente válidas para os grupos etários da população pesquisada, limitando-se aos indivíduos estudados.¹² Tal sugestão foi debatida em outras publicações.¹⁶⁻¹⁹ Portanto, sendo escolhida a amostragem por conglomerados, há um “preço” a ser pago pelo pesquisador por ter sua tarefa facilitada ao se investigar apenas os conglomerados sorteados.

Nesse sentido, propõe-se descrever uma amostragem complexa probabilística, por conglomerados em dois estágios, os métodos de estimação utilizados e a correção pelo efeito de desenho necessários na análise dos dados do Levantamento Epidemiológico das condições de Saúde Bucal da população de um município de porte médio “Projeto SBMOC”, cuja metodologia foi baseada no Projeto SBBrazil 2002/2003¹³⁻¹⁵. Portanto, com garantia da representatividade da população nas idades índices e faixas etárias preconizadas pela Organização Mundial de Saúde (OMS)²⁰ para investigação das condições de saúde bucal de uma população. Acredita-se que tal proposta possa contribuir para esclarecer e orientar outros pesquisadores sobre o tema, objeto de discussão em publicações prévias^{12,13,16-19}, apontado frequentemente como confuso e controverso.

METODOLOGIA

A definição do universo da pesquisa e das condições a serem investigadas foi feita

conforme preconiza a OMS que, em levantamentos epidemiológicos sobre condições de saúde bucal, deve-se garantir a representatividade da população nas idades índices de 5 e 12 anos e nas faixas etárias 15-19 anos (representando adolescentes), 35-44 anos (representando adultos) e 65-74 anos (representando idosos).²⁰ Dessa forma, foi definido que o “Projeto SBMOC” deveria ter uma amostra que garantisse representatividade para essas idades índices e faixas etárias, assim como para a faixa etária de 18-36 meses (representando bebês), conforme projeto SBBrazil 2002/2003.¹³⁻¹⁵ Garantiu-se poder de inferência e a proporcionalidade por sexo para as duas idades índices e cada um dos estratos etários. Foram estimadas seis diferentes amostras.

As condições pesquisadas e a seleção dos respectivos índices para avaliação dessas condições foram definidas para se orientar a elaboração do plano amostral. Os principais problemas pesquisados foram: cárie dentária, doença periodontal, má-oclusão, fluorose, uso e necessidades de prótese e lesões fundamentais em tecidos moles. Como seria desejável que a amostra tivesse poder de inferência para essas diferentes condições, ficou estabelecido que se buscasse o maior tamanho de amostra viável.

O cálculo do tamanho amostral foi feito considerando-se estimativas de proporções e de médias, a fim de compará-las e se optar pelo maior tamanho de amostra, exceto na faixa etária de 18 a 36 meses e na idade índice de 5 anos. Na estimativa que considerou proporções, optou-se por realizar os cálculos a partir da ocorrência dos eventos ou doenças em 50% da população, o que garante maior tamanho amostral.¹ Na estimativa para médias, definiu-se como parâmetro à cárie dentária, através do CPOD (índice que mede o ataque pela cárie dental através do número de dentes cariados, perdidos e obturados), apresentado no relatório final do Projeto SBBrazil 2002/2003, para a região sudeste¹⁵, nas duas idades índices e nas quatro faixas etárias preconizadas.

Nas duas propostas descritas anteriormente, utilizaram-se fórmulas que adotaram a correção para população finita, com o *deff* igual a 2,0 e uma taxa de não resposta de 20%, visto que qualquer estudo estará sujeito à perda de dados e o tamanho amostral é estabelecido a partir de um erro julgado pertinente, mas que será alterado se houver perda de dados (não resposta). Assim, para se resguardar a precisão, deve-se majorar um pouco a amostra a fim de se compensar eventuais perdas.¹ As fórmulas para cálculo do tamanho de amostra, para população finita quanto à proporção da presença da doença ou evento e a distribuição média²¹ utilizadas, estão apresentadas a seguir:

PROPORÇÃO

$$n = \frac{z^2 \cdot p \cdot q \cdot N}{\varepsilon^2 \cdot (N - 1) + z^2 \cdot p \cdot (1-p)} \cdot deff + 20\%$$

População finita

MÉDIA

$$n = \frac{z^2 \cdot s^2 \cdot N}{(\bar{x} \cdot \varepsilon)^2 \cdot (N - 1) + z^2 s^2} \cdot deff + 20\%$$

Onde:

n = tamanho da amostra;

z = 1,96 (valor tabelado da distribuição normal padronizada correspondente a um intervalo de 95%);

p = prevalência estimada da doença ou evento, que neste estudo foi de 50%;

q = 1- p;

ε = erro tolerável de amostragem, foi de 5,5% para proporção e 10% para a média;

N= tamanho da população;

s²= variância estimada do evento investigado, que nesse estudo foi do CPOD e ceod;

\bar{x} = média estimada do evento investigado, que nesse estudo foi o CPOD e ceod;

Deff = efeito de desenho.

A seleção da amostra refere-se à forma de seleção dos participantes. Para a coleta de dados, conduzida nos domicílios na área urbana, optou-se por uma amostra probabilística por conglomerados em dois estágios, sendo o primeiro constituído pelos setores censitários urbanos (Unidades Primárias de Amostragem - UPA) e o segundo constituído pelas quadras urbanas (Unidades Secundárias de Amostragem - USA). Na zona rural, optou-se por uma amostra probabilística por conglomerados em um único estágio, em que as unidades primárias de amostragem foram constituídas pelas áreas rurais.

Obteve-se o mapa do município com os setores censitários urbanos definidos pelo IBGE/2000. Constatou-se a existência de 276 setores urbanos.²² Como o município não possuía setores rurais²², utilizaram-se dados fornecidos pela vigilância epidemiológica do município a fim de se identificar áreas rurais. As unidades primárias de amostragem foram selecionadas de forma aleatória simples. No segundo estágio, um número de quadras foi sorteado em cada setor urbano, selecionado no primeiro estágio. Nessa etapa, foi planejada a exclusão das quadras urbanas não domiciliares (praças, igrejas, indústrias, hospitais, quartéis, presídios, entre outros). Na zona rural, todos os domicílios situados a uma distância de até 500 metros de uma instituição de referência foram selecionados.^{13,14} Todos os domicílios foram sequencialmente visitados e os indivíduos pertencentes às idades de interesse, convidados a participarem da pesquisa.

Na análise dos dados, obtidos em planos amostrais complexos, é necessário o uso de técnicas desenvolvidas especificamente para esse fim²³, ou seja, a ponderação pelo efeito de desenho. Na amostragem por conglomerados, a fim de se compensar as probabilidades (*f*) desiguais de seleção das unidades em cada um dos estágios (*f*₁; *f*₂;...*f*_i), devem-se atribuir ponderações diferenciadas aos elementos da amostra, correspondentes ao inverso do produto das probabilidades de inclusão nos diversos estágios de seleção, estimando, então, novos pesos (*w*) para cada elemento da amostra⁹:

$$w = 1/f$$

Onde: w = peso e f = probabilidade de inclusão

Inicialmente, foi calculada a probabilidade (f) de inclusão para, posteriormente, se chegar ao peso (w). Como na presente investigação haviam dois estágios, a probabilidade final seria obtida pelo produto da probabilidade de inclusão em cada um dos dois estágios ($f = f_1 \times f_2$). Como todos os domicílios, dentro das quadras eleitas, seriam potencialmente incluídos, eles não se constituíam como terceiro estágio de seleção. Entretanto, deve-se considerar a possibilidade de recusa na participação, o que acarretaria em diferentes possibilidades de inclusão. Assim, a taxa de resposta (t_{resposta}) em cada setor foi levada em consideração.⁹ Então:

$$f = f_1 \times f_2 \times t_{\text{resposta}}$$

Onde:

f_1 = probabilidade de inclusão no primeiro estágio (nº de setores sorteados / nº total de setores);

f_2 = probabilidade de inclusão no segundo estágio (nº de quadras sorteadas / nº total de quadras do setor);

t_{resposta} = Taxa de resposta.

Dessa forma, o peso final (w) pode ser assim definido:

$$w = \frac{1}{f_1 \times f_2 \times t_{\text{resposta}}}$$

Ressalta-se que f_1 foi constante (52/276), já f_2 e t_{resposta} variaram para cada setor.

Esse trabalho foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Estadual de Montes Claros (Unimontes) conforme parecer nº 318/2006, segundo Declaração de Helsinque e Resolução 196/96.

RESULTADOS

Tamanho da amostra: os cálculos evidenciaram a necessidade de se entrevistar e examinar 4478 indivíduos ao se considerar a prevalência dos eventos ou doenças em 50% da população, e a necessidade de se entrevistar e examinar 7551 indivíduos ao se considerar a média e desvio padrão do CPOD (Tabela 1).

Para cada idade índice e faixa etária, exceto para a faixa etária de 18 a 36 meses e idade de 5 anos, foi adotado o método que forneceu a maior estimativa do tamanho amostral. Dessa forma, obteve-se um tamanho amostral igual a 2359 para o sexo masculino e 2493 para o sexo feminino, totalizando 4852 indivíduos. A amostra coletada não foi igual à amostra planejada, segundo idade índice e faixa etária, como esperado. Observou-se que, para a idade índice de 12 anos, não foi atingido o tamanho amostral mínimo planejado. A proporção de taxa de não-reposta situou-se próxima a 10% (Tabela 2).

Foi estimada uma amostra probabilística por conglomerados em dois estágios. As unidades primárias de amostragem, referentes ao sorteio no primeiro estágio, foram constituídas de 52 setores censitários urbanos selecionados aleatoriamente entre os 276 setores definidos pelo IBGE para o censo demográfico de 2000 (Figura 1), e duas áreas rurais selecionadas entre as 11 áreas rurais do município.

No segundo estágio, em cada um dos 52 setores censitários urbanos selecionados, foram sorteadas, em média, sete quadras/setores (mínimo de 1 quadra e máximo de 11 quadras), correspondentes às unidades secundárias de amostragem. Foram sorteadas duas quadras em um setor, três em três setores, cinco em quatro setores, seis em quatro setores, sete em trinta e dois setores, oito em oito setores e onze em um setor. Portanto, no total, foram selecionadas 354 quadras, distribuídas entre os 52 setores.

Tabela 1: Tamanho das amostras por faixas etárias ou idades-índice, com proporcionalidade por sexo, quanto à prevalência e quanto à média de ceo-d e CPOD com correção pelo efeito de desenho (*deff*) e taxa de não resposta.

	População estimada *	Prevalência ou $\bar{x} \pm s$ estimados	Erro de amostragem	Tamanho da amostra parcial	Tamanho da amostra com correção pelo efeito de desenho ($deff = 2,0$)	Tamanho final da amostra com taxa de não resposta = 20%	Masculino		Feminino	
							População estimada*	Número de exames	População estimada*	Número de exames
							n (%)	n	n(%)	n
18-36 meses	20.360	0,5	0,055	314	628	754	10.180 (50)	377	10.180 (50)	377
5 anos	7.017	0,5	0,055	304	608	729	3.579 (51)	372	3.438 (49)	357
12 anos	7.767	0,5	0,055	305	610	732	3.883 (50)	366	3.884 (50)	366
15-19 anos	41.745	0,5	0,055	317	634	761	20.038 (48)	365	21.707 (52)	396
35-44 anos	46.471	0,5	0,055	318	636	762	22.306 (48)	366	24.435 (52)	396
65-74 anos	9.929	0,5	0,055	309	618	740	4.369 (44)	326	5.560 (56)	414
Total	133.289			1.867	3.732	4.478	64.355	2172	69.204	2.306

Quanto a Média (\bar{x}) e desvio-padrão (s) do ceod e CPOD

	N	$\bar{x} \pm s^{**}$	E	n	n	n	n (%)	n	n (%)	n
18-36 meses	20.360	0,95 \pm 2,33	0,10	1817	3633	4359	10.180 (50)	2180	10.180 (50)	2179
5 anos	7.017	2,50 \pm 3,36	0,10	561	1122	1346	3.579 (51)	687	3.438 (49)	659
12 anos	7.767	2,30 \pm 2,72	0,10	461	922	1106	3.883 (50)	553	3.884 (50)	553
15-19 anos	41.745	5,94 \pm 4,66	0,10	233	466	560	20.038 (48)	269	21.707 (52)	291
35-44 anos	46.471	20,30 \pm 7,74	0,10	56	112	134	22.306 (48)	64	24.435 (52)	70
65-74 anos	9.929	28,61 \pm 6,44	0,10	19	38	46	4.369 (44)	20	5.560 (56)	26
Total	133.289			3147	6293	7551	64.355	3773	69.204	3778

* Estimativa do IBGE para a população de Montes Claros, Minas Gerais em 2008¹⁵. ** Resultados encontrados no projeto SB Brasil 2002/2003 para região sudeste²².

Tabela 2: Tamanho da amostra planejado, amostra alcançada e taxa de resposta, com número de entrevistas e exames a serem realizados e nas idades-índice e faixas etárias quanto ao sexo.

Idades-índice Faixa etária	Tamanho da amostra planejado				Tamanho da amostra alcançado e taxa de resposta					
	População estimada*	Amostra estimada pela prevalência	Amostra estimada pelo ceod e CPOD	Amostra final	Amostra final masculina (masc.)	Amostra final feminina (fem.)	Total	Masc.	Fem.	Taxa de resposta
	N	N	n	n	n	N	n	n	n	%
18 a 36 meses	20.360	754	4.359	754	377	377	809	401	408	92,3
05 anos	7.017	729	1.346	729	372	357	997*	536	460	94,4
12 anos	7.767	732	1.106	1.106	553	553	357	165	192	92,8
15 a 19 anos	41.745	761	560	761	365	396	764	367	397	90,5
35 a 44 anos	46.471	762	134	762	366	396	846	383	463	91,4
65 a 74 anos	9929	740	46	740	326	414	736	313	423	92,0
Total	133.289	4.478	7.551	4.852	2.359	2.493	4.509	2.165	2.343	90,6

* Na idade de 5 anos, houve perda da informação de um participante quanto ao sexo.

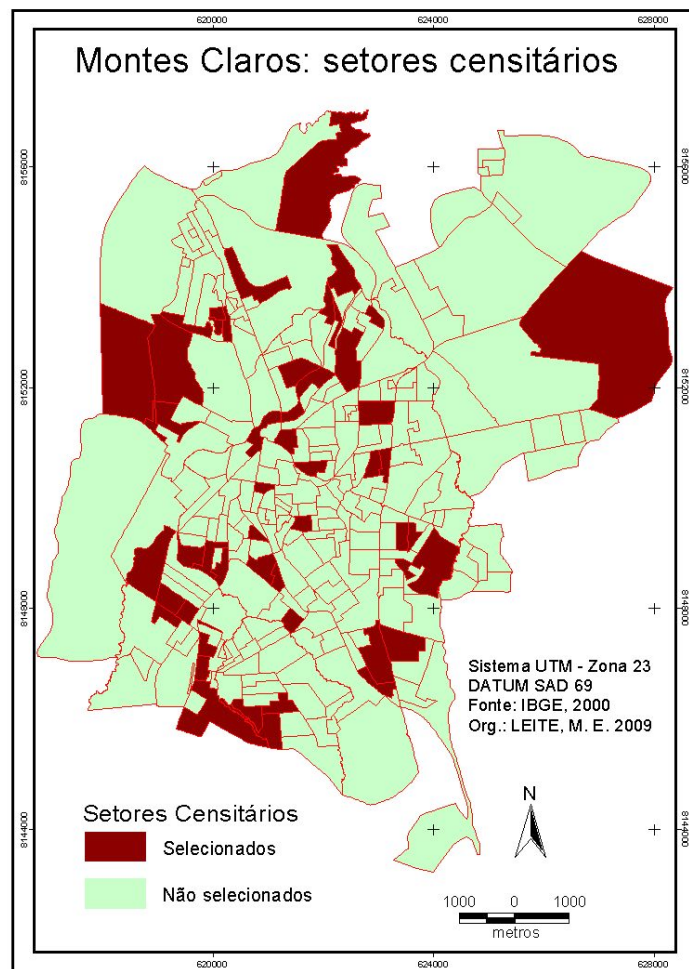


Figura 1: Esquema dos setores censitários urbanos de Montes Claros MG sorteados, conforme mapa do IBGE do censo demográfico de 2000.

Na ponderação pelo efeito de desenho, cada elemento da amostra teve associado um peso (w) que é o inverso da sua probabilidade de inclusão. Desse modo, construiu-se um conjunto de pesos cujos valores foram calculados em função da probabilidade (f_1) de inclusão do setor e da probabilidade de inclusão da quadra (f_2), onde se localiza o domicílio do elemento amostral. Nesse sistema de ponderação, o peso obtido para cada indivíduo variou de 3,7 a 47,1 apresentando média de 11,2 com desvio padrão igual a 7,9 (Tabela 3).

Na distribuição dos pesos amostrais por setor censitário, observou-se que cerca de 87% dos setores apresentaram pesos entre 3,7 a 20. Os valores mínimos e máximos da distribuição de frequência e a média e

desvio padrão dos pesos foram, respectivamente, 3,7 a 36,1 com média= $11,0 \pm 6,7$, na faixa etária de 18 a 36 meses; 3,7 a 47,1 com média= $11,6 \pm 8,3$, na idade índice de 5 anos; 3,6 a 47,5 com média= $11,4 \pm 8,4$, na idade índice de 12 anos; 3,7 a 36,1 com média= $11,4 \pm 7,0$, na faixa etária de 15 a 19 anos; 3,7 a 32,3 com média= $11,2 \pm 6,3$, na faixa etária de 35 a 44 anos; 3,6 a 36,2 com média= $11,1 \pm 6,9$, na faixa etária de 65 a 74 anos. No geral, o valor mínimo foi de 3,7 e o máximo de 47,1, média igual a 11,2 e desvio padrão $\pm 7,9$.

Após a estimativa da amostra, considerando os pesos amostrais por faixa etária ou idade índice, constataram-se correções na composição das amostras quanto à idade índice ou faixa etária (Tabela 4).

MARTINS, A. M. E. B. L.; SANTOS-NETO, P. E.; BATISTA, L. H. S.; NASCIMENTO, J. E.; GUSMÃO, A. F.; ELEUTÉRIO, N. B.; GUIMARÃES, A. L. S. G.; PAULA, A. M. B.; HAIKAL, D. S.; SILVEIRA, M. F.; PORDEUS, I. A.

Tabela 3: Distribuição dos pesos amostrais segundo setor censitário.

Peso	Nº de setores	%
3,7 a 9,9	31	59,7
10,0 a 19,9	14	26,9
20,0 a 29,9	6	11,5
30,0 a 39,9	0	0,0
40,0 a 47,1	1	1,9
Total	52	100,0

Tabela 4: Tamanho da amostra coletada com correção pelo efeito desenho.

Idades-índice Faixa etária	Amostra coletada (n)	n*
18 a 36 meses	809	8670,15
05 anos	997	13480,73
12 anos	357	5292,19
15 a 19 anos	764	7161,83
35 a 44 anos	846	8665,14
65 a 74 anos	736	7238,59

*Correção pelo efeito desenho.

DISCUSSÃO

Amostragens probabilísticas permitem inferências sobre a população da qual foram retiradas.⁶ No planejamento amostral, uma vez definidos os objetivos da pesquisa, a população a ser amostrada, bem como os parâmetros que precisam ser considerados para estimativa do tamanho da amostra, deve-se considerar a definição da unidade de amostragem e a forma de seleção dos elementos da população.⁵ No plano amostral da presente investigação, essas questões foram consideradas.²⁴ Portanto, o levantamento

epidemiológico das condições de saúde bucal da população do município investigado, no presente estudo, tem poder inferencial para todos os estratos avaliados. Ressalta-se entretanto, que por questões logísticas optou-se pelo menor tamanho de amostra na faixa etária de 18 a 36 meses e na idade índice de cinco anos. Assim, tais amostras são representativas para estimar a prevalência das condições bucais avaliadas. Outros levantamentos epidemiológicos, sobre condições de saúde bucal, conduziram o planejamento de uma amostragem probabilística por conglomerados^{25,24}, como o apresentado na presente investigação. Entretanto,

tais investigações não apresentaram, no planejamento ou nos resultados publicados, a ponderação necessária para correção pelo efeito de desenho para compensar as probabilidades desiguais de seleção dos elementos amostrais.¹²

As propostas da OMS, para o cálculo do tamanho da amostra com o intuito de conduzir levantamentos epidemiológicos em saúde bucal, têm sido modificadas.^{26,27} A proposta da OMS de 1977, mesmo trabalhando com um tamanho mínimo de amostra já predeterminado e sem levar em consideração o tamanho da população, era cautelosa, considerando o plano amostral uma etapa crítica.²⁶

Na edição de 1987²⁷, era preconizada metodologia de levantamento exploratório, o tamanho da amostra era um número padronizado em cada estrato a ser examinado e variava de 25 a 50, dependendo da prevalência e severidade da doença bucal no estrato. Nos estratos com prevalência estimada baixa ou muito baixa (20% ou mais livres de cárie), propunha-se o exame de 25 indivíduos, e, para moderada e alta prevalência (5 a 20% livres de cárie), propunha-se o exame de 40 a 50 pessoas. Ao contrário dos estudos de prevalência nos quais se avaliava um percentual da população (10% ou 1%), essa forma de obtenção da amostra preconiza o exame de 25 a 50 pessoas de cada estrato, independentemente do tamanho da população. Não foi apresentado o princípio utilizado para essa metodologia.^{28,29} Essas propostas tornaram-se incômodas para os especialistas que trabalhavam com planos amostrais complexos; afinal, uma das regras fundamentais de amostragem, a chamada lei dos grandes números, reza que “a aproximação relativa aumenta à medida que cresce o número de determinações”³⁰, reforçando, desse modo, a lógica da representatividade, ter uma relação direta com o tamanho da amostra, ou seja, quanto maior for a amostra, mais precisos são os resultados. Assim, 25 a 50 pessoas examinadas por estrato, independentemente de a cidade ter 20 mil ou 1 milhão de habitantes, era uma simplificação excessiva.

Os planos amostrais para levantamentos de

saúde bucal são, via de regra, complexos, dada a distribuição das doenças bucais. Segundo Moysés

[...] deve-se tomar cuidado, portanto, com o estabelecimento de uma amostra por critérios de decisão tais como [...] recorrer à amostragem de conveniência, quando as condições operacionais da pesquisa são desfavoráveis ou, ainda, lançar mão das conhecidas ‘amostras por números cabalísticos’, tais como as famosas 30 ou 385 unidades amostrais.³¹

O autor faz menção ao cálculo básico para determinação de tamanho da amostra, referido anteriormente, em que, ao se estimar uma prevalência de 50%, uma margem de erro de 5% e uma significância de 95%, o resultado é sempre 385. De certo modo, esse tipo de raciocínio (obtenção da amostra a partir da estimativa da prevalência) é impraticável para ser aplicado no caso da cárie dentária, por exemplo. O cálculo de tamanho da amostra, a partir da proporção, pressupõe que a unidade de medida da doença seja o indivíduo, ou seja, a prevalência é obtida pela contagem de indivíduos doentes, estabelecendo-se uma relação com a população total. Ocorre que, no caso das doenças bucais, a cárie dentária, por exemplo, a unidade em estudo é o dente, ou seja, não são contadas as pessoas atingidas, mas os dentes atingidos por pessoa. Assim, calcular uma amostra, a partir da proporção de indivíduos acometidos, serve apenas para determinar um número necessário para se estimar a prevalência de cárie por indivíduo. Sabe-se que esse é um dado irrelevante e sem o menor sentido, em função de a cárie dentária ser uma doença presente em mais de 90% da população. Somente em idades menores ou mesmo em localidades de baixíssima prevalência, este tipo de indicador adquire maior importância. Sendo o CPOD uma variável quantitativa, com variabilidade possível de ser estimada pelo desvio-padrão, o cálculo da amostra, a partir da média e do desvio padrão, é mais sensato.

MARTINS, A. M. E. B. L.; SANTOS-NETO, P. E.; BATISTA, L. H. S.; NASCIMENTO, J. E.; GUSMÃO, A. F.; ELEUTÉRIO, N. B.; GUIMARÃES, A. L. S. G.; PAULA, A. M. B.; HAIKAL, D. S.; SILVEIRA, M. F.; PORDEUS, I. A.

Em síntese, a determinação do tamanho da amostra em estudos de cárie dentária se constitui em um processo complexo, em que têm que ser levados em conta vários aspectos, entre eles os objetivos da investigação, as características da população, os instrumentos de medida, o nível de precisão pretendido (que está relacionado com o nível de planejamento das ações) e, também, a exequibilidade e o custo da investigação.³¹⁻³³

Outra questão a ser ressaltada é o fato de que a representatividade de uma amostra é dada, além da sua relação com a prevalência estimada da doença em estudo, pela aleatoriedade da escolha. Nesse caso, tão importante quanto se definir o tamanho da amostra é sua determinação do processo de sorteio. Nesse sentido, é de fundamental importância que, na medida do possível e resguardadas as peculiaridades do levantamento, todos os componentes da população tenham iguais chances de participarem da amostra. Isso se torna mais fundamental, ainda, quando se trabalha em populações com bases de dados pouco confiáveis, em que não se pode dispor de listas para proceder a aleatorização.²⁹ Observou-se, na presente investigação, uma distribuição desigual da população por faixa etária, conforme esperado, uma vez que quando se compara uma amostra aleatória simples de n elementos a uma amostra por conglomerados (a conglomerados contendo n elementos), verifica-se que o número de seleções independentes se reduz de n para a , aumentando a variância do estimador de proporção, dado que quase invariavelmente as classes de interesse dentro dos conglomerados não têm distribuição independente entre si.³⁴ Sabe-se que a dependência das observações, devido ao processo de amostragem por conglomerado, deve ser sempre levada em consideração na análise de grandes levantamentos de saúde.³⁵

Sendo, assim, considerando que o objetivo principal do levantamento epidemiológico descrito no presente manuscrito foi estimar prevalência das variáveis relacionadas à saúde ou distribuições médias da cárie na população investigada, a calibração dos pesos naturais foi necessária para lidar com os problemas de coleta e seleção. Ressalta-se portanto que o desenho

amostral, isto é, de que forma as unidades amostrais são selecionadas, exerce influência no tamanho amostral.³⁶ Do ponto de vista prático (ou operacional), uma amostragem por conglomerados é mais simples e, frequentemente, a mais usada.³⁷

Mas, ao se investigar apenas os *clusters* sorteados há um aumento da imprecisão devido às possíveis correlações das unidades amostrais dentro e entre os *clusters*, daí a necessidade do uso do *deff*.¹ O *deff* ou efeito do plano amostral representa a razão entre a variância de uma estimativa obtida com uma amostra de conglomerados e a variância que seria obtida com uma amostra aleatória simples de mesmo tamanho.¹ O efeito do desenho é um importante indicador para avaliação do erro de amostragem o qual permite avaliar subestimativas ou mesmo superestimativas dos erros padrão, utilizando-se as diferentes características do plano de amostragem e diferentes métodos de estimação.³⁶

Para se atribuir um valor razoável ao *deff*, deveria se ter uma ideia da variabilidade das observações entre e dentro dos *clusters*, o que, frequentemente, não se dispõe. No caso de um estudo de prevalência, quanto mais próximas as prevalências entre os *clusters*, menor seria o *deff* necessário. Na prática, salvo em situações especiais, um *deff* de 1,4 ou 1,5 (uma correção no tamanho amostral entre 40% e 50%) já deve ser suficiente para resguardar a precisão desejada do pesquisador.¹ No presente estudo, o efeito desenho utilizado foi de 2,0. Embora o cálculo dos efeitos de desenho, após a coleta de dados, seja recomendado para saber se o tamanho amostral utilizado foi suficiente para atender as especificações de precisão previamente estabelecidas,³⁶ quando se deseja conhecer a condição de saúde bucal em uma localidade, tal estimativa não foi apresentada na presente investigação uma vez que foram feitos, na verdade, seis planejamentos amostrais e foram propostas avaliações de diversas condições de saúde, tornando-se inviáveis tais cálculos. Por outro lado, a ponderação conduzida no presente estudo teve impacto sobre as estimativas.

CONCLUSÕES

A amostragem complexa probabilística por conglomerados em dois estágios e os métodos de estimação utilizados, assim como a correção pelo efeito de desenho necessário na análise dos dados de um Levantamento Epidemiológico, foram baseados nas propostas mais recentes de amostragem. Sendo assim, a amostra do inquérito em questão apresenta um poder de inferência necessário às investigações epidemiológicas.

Fonte de financiamento da pesquisa: Fundação de Amparo a Pesquisa do Estado de Minas Gerais – FAPEMIG Número do processo EDT 3270/06

Os autores declaram a inexistência de conflitos de interesses.

AGRADECIMENTOS

Agradecemos o apoio logístico da Unimontes e da Prefeitura Municipal de Montes Claros, o fomento da Fundação de Amparo a Pesquisa Estado de Minas Gerais FAPEMIG e a colaboração dos participantes. As pesquisadoras Andréa Maria Eleutério de Barros Lima Martins, Desirée Sant’Ana Haikal e Marise Fagundes da Silveira receberam bolsa da FAPEMIG. A Pesquisadora Isabela Almeida Pordeus é Bolsista de Produtividade em Pesquisa 1B do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPQ.

REFERÊNCIAS

1. LUIZ, R. R.; MAGNANINI, M. M. F. A lógica da determinação do tamanho da amostra em investigações epidemiológicas. *Cadernos de Saúde Coletiva*, Rio de Janeiro, v.8, n.2, p. 9-28, 2000.
2. BATTIST, I. D. E. *Análise de dados epidemiológicos incorporando planos amostrais complexos*. Tese - [Porto Alegre (RS)]: Universidade Federal do Rio Grande do Sul; 2008. Disponível em <<http://hdl.handle.net/10183/14675>>. Acesso em março de 2011.
3. BUSSAB, W. O.; MORETTIN, P. *Estatística básica*. 4 ed. São Paulo (SP): Atlas, 1987.
4. MAROTTI, J. *et al.* Amostragem em Pesquisa Clínica : tamanho da amostra. *Revista de Odontologia da Universidade Cidade de São Paulo*, São Paulo, v. 20, n. 2, p. 186-194, 2008.
5. BARBETTA, P. A. *Estatística aplicada às Ciências Sociais*. 4 ed. Florianópolis (SC): Ed. da UFSC, 2001.
6. FONSECA, J. S.; MARTINS, G. A. *Curso de estatística*. 6 ed. São Paulo (SP): Atlas, 1996.
7. SOUZA, M. H.; SILVA, N. N. Estimativas obtidas de um levantamento complexo. *Revista de Saúde Pública*, São Paulo, v. 37, p. 662-70, 2003.
8. CORDEIRO, R. Efeito do desenho em amostragem de conglomerado para estimar a distribuição de ocupações entre trabalhadores. *Revista de Saúde Pública*, São Paulo, v. 35, n. 1, fev, 2001.
9. SZWARCOWALD CL, DAMACENA GN. Amostras complexas em inquéritos populacionais: planejamento e implicações na análise estatística dos dados. *Revista Brasileira de Epidemiologia*, São Paulo, v. 11, n. 1, p. 38-45, 2008.
10. KORN, E. L.; GRAUBARD, B. I. Epidemiologic studies utilizing surveys: accounting for the sampling design. *American Journal of Public Health*, v. 81, n. 9, p. 1166-73, 1991.
11. SILVA, P. L. N.; PESSOA, D. G. C. P.; LILA, M. F. Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral. *Ciência & Saúde Coletiva*, v. 7, n. 4, p. 659-670, 2002.
12. QUEIROZ, R. C. S.; PORTELA, M. C.; VASCONCELLOS, M. T. L. Pesquisa sobre as Condições de Saúde Bucal da População Brasileira (SB Brasil 2003): seus dados não produzem estimativas populacionais, mas há possibilidade de correção. *Cadernos de Saúde Pública*, v. 25, n.1, p. 47-58, jan., 2009.

MARTINS, A. M. E. B. L.; SANTOS-NETO, P. E.; BATISTA, L. H. S.; NASCIMENTO, J. E.; GUSMÃO, A. F.; ELEUTÉRIO, N. B.; GUIMARÃES, A. L. S. G.; PAULA, A. M. B.; HAIKAL, D. S.; SILVEIRA, M. F.; PORDEUS, I. A.

13. RONCALLI, A.G. *et al.* Projeto SB2000: uma perspectiva para a consolidação da Epidemiologia em Saúde Bucal Coletiva. *Revista Brasileira de Odontopediatria e Saúde Coletiva*, v. 1, n. 2, p. 9-25, 2000.
14. BRASIL. Ministério da Saúde. Secretaria de Políticas de Saúde. Departamento de Atenção Básica. Área Técnica de Saúde Bucal. *Projeto SB2000: condições de saúde bucal da população brasileira no ano 2000: manual do examinador / Secretaria Políticas de Saúde, Departamento de Atenção Básica, Área Técnica de Saúde Bucal.* – Brasília: Ministério da Saúde, 2001.
15. BRASIL. Departamento de Atenção Básica, Secretaria de Atenção à Saúde, Ministério da Saúde. Projeto SB Brasil 2003. *Condições de saúde bucal da população brasileira, 2002-2003: resultados principais.* Brasília: Ministério da Saúde; 2004. Disponível em <http://www.apcd.org.br/prevencao/arquivos/projeto_sb_brasil.pdf> Acesso em: abr. 2009.
16. NARVAI, P. C. *et al.* Validade científica de conhecimento epidemiológico gerado com base no estudo Saúde Bucal Brasil 2003. *Cadernos de Saúde Pública*, v. 26, n. 4, p. 647-57, abr., 2010. Disponível em: <<http://www.scielo.br>>. Acesso em: 14 jul. 2011.
17. PORTELA, M. C.; VASCONCELLOS, M. T. L. Debate sobre o artigo de Narvai et al. *Cadernos de Saúde Pública*, v. 26, n. 4, p. 658-60, abr., 2010. Disponível em:<<http://www.scielo.br>>. Acesso em: 14 jul. 2011.
18. CASTIEL, L. D. Conflitos, interesses e alegorias: o caso SB Brasil 2003. *Cadernos de Saúde Pública*, v. 26, n. 4, p. 660-62, abr., 2010. Disponível em:<<http://www.scielo.br>>. Acesso em: 14 jul. 2011.
19. MOREIRA, R. S. Validade ou verdade? Algumas reflexões sobre crenças verdadeiras e justificáveis em epidemiologia. *Cadernos de Saúde Pública*, v. 26, n. 4, p. 663-65, abr., 2010. Disponível em: <<http://www.scielo.org/scielo>>. Acesso em: 14 jul. 2011.
20. WORLD HEALTH ORGANIZATION. *Oral Health surveys: basic methods.* 4. ed. Geneva: ORH EPID, 1997.
21. TRIOLA, M. F. *Introdução à Estatística.* 7 ed. Rio de Janeiro (RJ): LTC, 1999.
22. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. *Manual do recenseador - CD1.09.* Rio de Janeiro (RJ): IBGE, 2000. 151p.
23. BUSSAB, W.O., grupo de estudos em população, sexualidade e Aids. Plano amostral da Pesquisa Nacional sobre Comportamento Sexual e Percepções sobre HIV/ Aids, 2005. *Revista de Saúde Pública* [online], n. 42, suppl 1.1: p. 12-20, 2008. Disponível em: <<http://www.scielo.br/pdf/rsp/v42s1/04.pdf>> Acesso em: jun. 2011.
24. FRIAS, A. C.; ANTUNES, J. L. F.; NARVAI, P.C. Precisão e validade de levantamentos epidemiológicos em saúde bucal: cárie dentária na cidade de São Paulo, 2002. *Revista Brasileira de Epidemiologia*, v. 7, n. 2, p. 144-54, 2004.
25. BRASIL. SB Brasil 2010. *Manual de Calibração de Examinadores.* Brasília: Ministério da Saúde, 2010. 21p.
26. WORLD HEALTH ORGANIZATION. *Oral health surveys: basic methods.* 2. ed. Geneva: ORH/EPID, 1977.
27. WORLD HEALTH ORGANIZATION. *Oral health surveys: basic methods.* 3. ed. Geneva: ORH/EPID, 1987.
28. WORLD HEALTH ORGANIZATION. *Calibration of Examiners for Oral Health Epidemiology Surveys.*

Technical Report. Geneva: ORH/EPID, 1993.

29. OLIVEIRA, A. G. R. C. *et al.* Levantamentos epidemiológicos em saúde bucal: análise da metodologia proposta pela Organização Mundial da Saúde. *Revista Brasileira de Epidemiologia*, v. 1, n. 2, p. 177-89, ago., 1998. Disponível em: <<http://www.scielosp.org>>. Acesso em: 21 jul. 2011.

30. SOUNIS, E. *Manual de Higiene e Medicina do Trabalho*. São Paulo (SP): Mc Graw- Hill do Brasil Ltda, 1975.

31. MOYSÉS, S. J. O conceito de promoção da saúde na construção de sistemas de atenção em saúde bucal coletiva. In: KRIGER, L., coordenador. *Promoção de saúde bucal*. São Paulo (SP): Artes Médicas, 1997. p. 373-407.

32. MANJI, F.; FEJERSKOV, O. Um enfoque epidemiológico para a cárie dentária. In: THYLSTRUP, A.; FEJERSKOV, O. *Cariologia clínica*. São Paulo (SP): Editora Santos, 1995. Cap. 7, p. 159-92.

33. ORGANIZAÇÃO MUNDIAL DA SAÚDE. *Levantamento epidemiológico básico de saúde bucal:*

manual de instruções. 3. ed. São Paulo (SP): Editora Santos, 1991.

34. KISH, L. *Muestreo de encuestas*. Mexico (DF): Editorial Trillas; 1972.

35. KORN, E. L.; GRAUBARD, B. I. Analysis of large health surveys: accounting for the sampling design. *Journal of the Royal Statistical Society*, v. 158, n. 2, p. 263-95, 1995. Disponível em: <<http://www.jstor.org/stable/2983292>>. Acesso em: mai 2011.

36. SILVA, N. N. *Amostragem probabilística: um curso introdutório*. São Paulo (SP): EDUSP, 1998.

37. LEVY, P. S.; LEMESHOW, S. *Sampling of Populations: methods and applications*. New York: John Wiley & Sons, 1999.

Autor para correspondência:

Andréa Maria Eleutério de Barros Lima Martins
Universidade Estadual de Montes Claros
Campus Universitário Professor Darcy Ribeiro
Vila Mauricéia - Montes Claros - Minas Gerais - Brasil
martins.andreambl@gmail.com