

## Planejamento amostral para reduzir a impressão na amostragem por conglomerados em duas etapas

Mariano M. Espinosa<sup>1</sup>, Iury A. Rodrigues<sup>2</sup>, Jackson Pereira Júnior<sup>3†</sup>, Rafael L. S. Maciel<sup>4</sup>

<sup>1</sup>Professor, Departamento de Estatística, UFMT. E-mail: [marianom@cdp.ufmt.br](mailto:marianom@cdp.ufmt.br).

<sup>2</sup>Graduando em Estatística, UFMT. E-mail: [iuryrodrigues60@gmail.com](mailto:iuryrodrigues60@gmail.com).

<sup>3</sup>Professor, IFMT.

<sup>4</sup>Graduando em Estatística, UFMT. E-mail: [maciel.raf@gmail.com](mailto:maciel.raf@gmail.com).

**Resumo:** O objetivo do presente trabalho foi apresentar um planejamento amostral para reduzir a imprecisão na amostragem por conglomerados em duas etapas. Trata-se de um estudo metodológico sobre o planejamento amostral e determinação do tamanho da amostra em pesquisas de base populacional. Uma vez que neste tipo de pesquisas científicas em geral é utilizada amostragem por conglomeradas em duas ou mais etapas para inferir sobre uma determinada população. No entanto, ao se realizar uma pesquisa com conglomerados sorteados aleatoriamente existe um aumento da imprecisão devido às possíveis correlações das unidades amostrais dentro e entre os conglomerados, por esse motivo quando conglomerados são utilizados, usualmente é considerada uma correção denominada de efeito de desenho. Assim para determinar o tamanho de amostra utiliza-se a amostragem aleatória simples e a seguir este valor é corrigido multiplicando-o pelo valor do efeito do desenho. Além disso é importante estratificar a população para reduzir a variância e ainda considerar um acréscimo para possíveis ausência de resposta. Para isso no presente estudo o tamanho da amostra foi multiplicado pelo efeito do desenho e adicionado uma taxa mínima de resposta estimada igual a 85%. Com o procedimento apresentado neste estudo espera-se uma boa representatividade na amostra, por conglomerados em dois estágios em cada município considerado na pesquisa. Considera-se que este estudo seja de grande relevância para o planejamento amostral das diversas pesquisas que pretendam realizar estudos de base populacional.

**Palavras chave:** população, amostra, amostragem, estratos, conglomerado

**Abstract:** The objective of this work is to present a sampling plan to reduce imprecision in two-stage cluster sampling. This is a methodological study on sample planning and sample size determination in population-based surveys. Since in this type of scientific research in general is used cluster sampling in two or more steps to infer about a certain population. However, when conducting a randomly randomized cluster survey, there is an increase in imprecision because of the possible correlations of sample units within and between clusters. For this reason, when clusters are used, it is usually considered a design effect correction. Thus, to determine the sample size, simple random sampling is used and then this value is corrected by multiplying it by the value of the drawing effect. In addition, it is important to stratify the population to reduce variance and to consider an addition for possible non-response. For this in the present study the sample size was multiplied by the design effect and an estimated minimum response rate of 85% was added. With the procedure presented in this study it is expected a good representativeness in the sample, by conglomerates in two stages in each municipality considered in the research. It is considered that this study is of great relevance for the sample planning of several studies that intend to carry out population-based studies.

**Keywords:** population, sample, sampling, estrates, cluster.

### Introdução

Os dados desta pesquisa foram planejados para serem coletados por meio de uma amostra probabilística de domicílios, por conglomerados, em dois estágios de seleção: as unidades primárias constituídas pelos setores censitários e as unidades secundárias pelos domicílios.

---

† Autor correspondente: [jackson.jrp@gmail.com](mailto:jackson.jrp@gmail.com).

Para maior representatividade da população dos municípios estudados, inicialmente, foi realizada uma amostragem aleatória simples. A seguir foi realizado um sorteio em dois estágios. O sorteio com probabilidade proporcional ao tamanho dos conglomerados é um processo que controla o tamanho da amostra de elementos, além de mantê-la autoponderada. Nestes estágios, o método utilizado para o sorteio também foi por amostragem probabilística. Fundamentado nos princípios básicos de eficiência, é amplamente aceito e usado em inquéritos (KIS, 1965 e SILVA, 2001). Portanto, o objetivo deste estudo é apresentar um planejamento amostral para reduzir a imprecisão na amostragem por conglomerados em duas etapas.

## Metodologia

Para definir os setores por município foi utilizado o sistema de delimitação de setores do IBGE (2010) segundo a classificação de setores e o número de domicílios existentes em cada setor do município.

Para mostrar a aplicação deste estudo foi considerada a população constituída pelos moradores dos municípios de Alta Floresta, Diamantino, Sinop e Sorriso da área urbana, no Estado de Mato Grosso, pertencentes à área de influência da BR 163. Estes municípios foram selecionados por possuírem escritórios regionais de saúde, exceto Sorriso que foi escolhido por proximidade geográfica do município de Sinop, além de questões políticas e de apoio logístico.

A seguir são apresentadas as populações e domicílios de cada município (Tabela 1 e 2).

Tabela 1 População total por município para a área urbana de quatro municípios do Estado de Mato Grosso.

Município	Alta Floresta	Diamantino	Sinop	Sorriso	Total
Total	40091	15895	93696	58363	208045

Fonte: IBGE - Censos Demográficos e Contagem Populacional; para os anos intercensitários, estimativas preliminares dos totais populacionais, estratificadas por idade e sexo pelo MS/SE/Datasus, 2010.

Tabela 2 Domicílios\* por município para a área urbana dos quatro municípios do Estado de Mato Grosso considerados.

Município	Alta Floresta	Diamantino	Sinop	Sorriso	Total
Total	14003	5360	31165	18807	69335

Fonte: IBGE - Censos Demográficos e Contagem Populacional; para os anos intercensitários, estimativas preliminares dos totais populacionais, estratificadas por idade e sexo pelo MS/SE/Datasus, 2010. \*: Domicílios particulares permanentes.

Nesta pesquisa o planejamento de amostragem foi realizado pelo tipo de amostragem probabilístico, considerando os métodos de amostragem aleatória simples e por conglomerado em dois estágios (BOLFARINE & BUSSAB, 2005; CESAR, et al., 2005). O primeiro estágio é constituído pelos setores censitários urbanos (Unidades Primárias de Amostragem - UPA) e o segundo constituído pelos domicílios (Unidades Secundárias de Amostragem - USA). Uma vez que, somente se investigam os conglomerados sorteados, pode existir aumento da imprecisão devido às possíveis correlações das unidades amostrais dentro e entre os conglomerados, fazendo-se necessário aumentar o tamanho da amostra. Assim quando conglomerados são utilizados, usualmente é considerada uma correção denominada de efeito de desenho (*design effect*), que teoricamente é a razão entre a variância estimada pelo método de amostragem por conglomerado dividido pela variância estimada pelo método de amostragem aleatória simples (Silva, 2001 e Bolfarine e Bussab, 2005). Assim para determinar o tamanho de amostra utiliza-se a amostragem aleatória simples e a seguir este valor é corrigido multiplicando-o pelo valor do efeito do desenho. Segundo Medronho et

**Sigmae**, Alfenas, v.8, n.2, p. 698-703, 2019.

64ª Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).

18º Simpósio de Estatística Aplicada à Experimentação Agrônômica (SEAGRO).

al (2002) na prática um efeito de desenho de 1,4 ou 1,5, que corresponde a uma correção no tamanho de amostra entre 40% e 50% deve ser suficiente para resguardar a precisão desejada pelo pesquisador.

Para determinar o tamanho de amostra aproximado ( $n'$ ), necessário para estimar  $p$ , foi utilizada a seguinte expressão (SCHEAFFER, et al, 1987 e LEVY, P.S. & LEMESHOW, 2008):

$$n' = \frac{Np(1-p)}{(N-1)(d/z_{\alpha/2})^2 + p(1-p)} \quad (1)$$

Ao utilizar conglomerados para determinar o tamanho de amostra, o valor obtido pela expressão (1) foi multiplicado pelo efeito de desenho ( $deff$ ) e efetuando um acréscimo para a ausência de resposta, considerando uma taxa mínima de resposta estimada igual a 85%. Assim o tamanho de amostra é dado pela expressão (2):

$$n = n' \times deff + (n' / 0,85 - n') \quad (2)$$

- $n$  = Tamanho de amostra ou número de unidades na amostra;
- $N$  = Tamanho da população ou número de unidades na população;
- $N$  = Número de unidades na população em cada município;
- $p$  = Proporção populacional;
- $d$  = Limite para o erro de estimação ou erro tolerável de amostragem;
- $Z_{\alpha/2}$  = Obtido da tabela da distribuição normal padronizada, o qual representa o nível de confiança;
- $deff$  = Efeito do desenho.

## Resultados e discussão

Utilizando-se a Expressão (2) e as informações da Tabela 1 foi determinado o tamanho da amostra na população estudada, considerando o nível de confiança de 95%, isto é,  $z_{\alpha/2} = 1,96$ , uma proporção de 0,5% ( $p=0,5$ ), um erro de estimação de 2,5% ( $d=0,025$ ) e um efeito de desenho de 1,5 ( $deff=1,5$ ). Assim o tamanho de amostra da população considerada coregida e com acréscimo foi de aproximadamente 2557 indivíduos. Também foi determinado o tamanho da mostra por município ( $n_i$ ) proporcional ao tamanho da população estes valores são apresentados na Tabela 3.

Tabela 3 Tamanho da população e amostra por município para a área urbana dos quatro municípios do Estado de Mato Grosso considerados, Cuiabá MT, 2019.

Município	Alta Floresta	Diamantino	Sinop	Sorriso	Total
$N_i$	40091	15895	93696	58363	208045
$n_i$	493	195	1152	717	2557

Fonte: IBGE - Censos Demográficos e Contagem Populacional; para os anos intercensitários, estimativas preliminares dos totais populacionais, estratificadas por idade e sexo pelo MS/SE/Datusus, 2010.

Na Tabela 4 é apresentado o número de setores e média de moradores por domicílio.

**Sigmae**, Alfenas, v.8, n.2, p. 698-703, 2019.

64ª Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).

18º Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO).

Tabela 4 Número de setores e média de moradores por domicílio dos municípios em estudo.

Município	Número de Setores	Média de moradores por Domicílio
Alta Floresta	64	2,86
Diamantino	27	2,97
Sinop	183	3,01
Sorriso	93	3,10
Total	367	3,00

Fonte: IBGE - Censos Demográficos e Contagem Populacional; para os anos intercensitários, estimativas preliminares dos totais populacionais, estratificadas por idade e sexo pelo MS/SE/Datasus, 2010.

Considerando que os indivíduos são identificados no espaço geográfico por meio de seus domicílios, para cada município foi determinado o número de domicílios ( $D_a$ ) a serem sorteados na pesquisa. Para tanto, o número de domicílios foi determinado dividindo-se o número de indivíduos da amostra pelo número médio de moradores por domicílio, utilizando as Tabelas 3 e 4. O número total de domicílios, por município, a ser considerado na pesquisa são apresentados na Tabela 5.

Tabela 5 Número total de moradores e domicílios a ser considerado na pesquisa, por município.

Municípios	Número de moradores (amostra)	Média de moradores por domicílio	O número total de domicílios por município ( $D_a$ )
Alta Floresta	493	2,86	172
Diamantino	195	2,97	66
Sinop	1152	3,01	383
Sorriso	717	3,10	231
Total	2557	3,00	852

De posse do total de domicílios listados e do número de domicílios a serem efetivamente selecionados, por setor em cada município. Para definir os setores por município, foi utilizado o sistema de delimitação de setores do IBGE (2010) segundo a classificação de setores e o número de domicílios existentes em cada município, apresentada na Tabela 6.

Do total dos setores censitários por município e da classificação de setores, considerou-se, no presente estudo aproximadamente 25% de cada município, totalizando 91 setores, para a determinação do número de setores censitários a serem sorteados (Tabela 7).

Tabela 6. Distribuição dos setores censitários existentes por município, segundo classificação do IBGE (2010).

Classificação dos setores*	Número de setores censitários por municípios				Total
	Alta Floresta	Diamantino	Sinop	Sorriso	
$\geq 200$ domicílios	39	13	73	47	172
$100 \leq \text{domicílios} < 200$	24	10	65	31	130
$< 100$ domicílios	1	4	45	15	65
Total	64	27	183	93	367

Fonte: IBGE (2010). \*: IBGE (2010).

Tabela 7. Distribuição dos setores censitários a serem sorteados por município, segundo classificação do IBGE (2010).

Classificação dos setores*	Número de setores censitários por municípios				Total
	Alta Floresta	Diamantino	Sinop	Sorriso	
≥200 domicílios	10	3	18	12	43
100≤domicílios<200	6	2	16	8	32
<100 domicílios	0	1	11	4	16
Total	16	6	45	24	91

Fonte: Os autores. \*: IBGE (2010).

Assim, com os números de setores apresentados na Tabela 7, realizou-se o sorteio aleatório dos mesmos, que corresponde ao primeiro estágio (UPA), em que os conglomerados são os setores censitários. Este sorteio foi realizado utilizando-se o pacote estatístico MINITAB versão 17.0.

Após o sorteio do número de setores censitários para o estudo procedeu-se à listagem de todos os domicílios pertencentes a esses setores sorteados, considerando o cadastramento do IBGE (2010), nos quatros municípios.

Com a distribuição dos setores a serem sorteados por município e o total de domicílios a ser considerado na pesquisa por município, foi possível definir o número de domicílios aproximados por classificação dos setores ( $ndapcs$ ) de acordo com o número de domicílios por setor e por município. Para isso o número total de domicílios por município foi dividido pelo número de setores censitários classificados por município ( $nsccpm$ ):

$$ndapcs = \frac{D_a}{nsccpm} \quad (3)$$

Por exemplo, em alta floresta o número de domicílios por setor será de aproximadamente 11 domicílios por setor. Isto é,

$$ndapcs = \frac{D_a}{nsccpm} = \frac{172}{16} \approx 11 \text{ domicílios por setor no município.}$$

A relação com todos os domicílios por municípios e por classificação dos setores a serem sorteados são apresentados na Tabela 8.

Assim, com os números apresentados na Tabela 8 será efetuado o sorteio aleatório dos domicílios por setores em cada município, que corresponde ao segundo estágio (USA), onde os conglomerados são os domicílios em cada setor sorteado. Este sorteio também será realizado utilizando um pacote estatístico por exemplo o MINITAB versão 17.0.

Tabela 8. Distribuição dos setores censitários a serem sorteados por município, segundo classificação do IBGE (2010).

Classificação dos setores*	Número de setores censitários por municípios				Total
	Alta Floresta	Diamantino	Sinop	Sorriso	
≥200 domicílios	110	33	162	120	425
100≤domicílios<200	66	22	144	80	312
<100 domicílios	0	11	99	40	150
Total	176	66	405	240	887

Fonte: Os autotres. \*: IBGE (2010).

**Sigmae**, Alfenas, v.8, n.2, p. 698-703, 2019.

64ª Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).

18º Simpósio de Estatística Aplicada à Experimentação Agrônômica (SEAGRO).

## Conclusão

Ao se realizar uma pesquisa com conglomerados sorteados aleatoriamente existe um aumento da imprecisão devido às possíveis correlações das unidades amostrais dentro e entre os conglomerados, por esse motivo a necessidade de utilizar o efeito do desenho no planejamento amostral. Assim para determinar o tamanho de amostra foi utilizado a amostragem aleatória simples e a seguir este valor foi corrigido multiplicando-o pelo valor do efeito do desenho. Na prática um efeito de desenho de 1,4 ou 1,5, que corresponde a uma correção no tamanho de amostra entre 40% e 50% deve ser suficiente para resguardar a precisão desejada pelo pesquisador. Também é importante estratificar a população para reduzir a variância e ainda considerar um acréscimo para possíveis ausência de resposta, para isso foi utilizada a expressão (2). Assim espera-se uma boa representatividade na amostra, por conglomerados em dois estágios em cada município considerado na pesquisa, embora não tenha sido utilizado o procedimento de uma amostragem complexa. Considera-se que este estudo seja de grande relevância para o planejamento amostral das diversas pesquisas que pretendam realizar estudos de base populacional.

## Referências bibliográficas

BOLFARINE, H.; BUSSAB, W. O. *Elementos de amostragem*. Editora Edgar Blücher. São Paulo, 2005, 274p.

CESAR, C.L.G.; CARANDINA, L., ALVES, M.C.G.P; BAROS, M.B.; GOLDBAUM, M. Inquérito multicêntrico de Saúde no Estado de São Paulo. ISA-SP. São Paulo: USP/FSP, 2005.

IBGE. *Censos Demográficos e Contagem Populacional; para os anos intercensitários, estimas preliminares dos totais populacionais, estratificado por idade e sexo pelo MS/SE/DATASUS*. Mato Grosso, 2005.

KISH, L. *Survey sampling*. John Wiley & Sons, inc., New York, 1965, 664p.

LEVY, P.S.; LEMESHOW, S. *Sampling of populations. Methods and Applications*. John Wiley & Sons, Inc. New York, 2008, 616p.

SCHEAFFER, R.L.; MENDENHALL, W. e OTT, L. *Elementos de Muestreo*. Grupo editorial Iberoamericana S. A. México, 1987, 321p.

SILVA, N.N. *Amostragem probabilística: um curso introdutório*. São Paulo: EDUSP; 2001, 120p.

**Sigmae**, Alfenas, v.8, n.2, p. 698-703, 2019.

64ª Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).  
18º Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO).