

Plano Amostral para a Pesquisa por Amostra de Domicílios em Minas Gerais – PAD-MG

Emilio Suyama

1. Introdução

Este relatório descreve o plano amostral proposto para a Pesquisa por Amostra de Domicílios em Minas Gerais, que será realizada pela Fundação João Pinheiro, a partir de 2009. O objetivo da pesquisa é fornecer uma ampla base de dados sobre vários aspectos sócio-econômicos da população mineira, como: educação, saúde, trabalho, renda, consciência ambiental, gastos individuais e do domicílio, informações sobre a juventude e protagonismo juvenil, para subsidiar a definição e avaliação anual de Políticas Públicas no Estado de Minas Gerais.

Os resultados devem ser apresentados segundo a estratificação geográfica, tanto do ponto de vista das Regiões de Planejamento, quanto das Mesorregiões, sendo utilizada alocação **potência** em cada parte da população, resultante do duplo critério de estratificação. Há interesse também nos resultados desagregados, para a cidade de Belo Horizonte, e para a Região Metropolitana de Belo Horizonte, que pertencem tanto à Região de Planejamento Central, quanto à Mesorregião Metropolitana. Outro interesse recai nos resultados agregados tanto para o Grande Norte, que contempla as Regiões de Planejamento: Norte, Jequitinhonha/Mucuri e Doce, quanto para o Estado de Minas Gerais, para as situações: urbana e rural.

2. Plano Amostral

2.1 População e sua estratificação

O Estado de Minas Gerais é constituído por 10 Regiões de Planejamento ou por 12 Mesorregiões. A população de interesse para a pesquisa é constituída pelos domicílios mineiros, tanto em situação urbana quanto na rural. A base de dados é composta pela listagem dos setores censitários feita pelo Censo Demográfico de 2000, com a eliminação de setores compostos apenas de domicílios coletivos, e que foram

definidos pelo IBGE como sendo quartéis, bases militares, alojamentos, acampamentos, embarcações, aldeias indígenas, penitenciárias, colônias penais, presídios, cadeias, asilos, orfanatos, conventos e hospitais. A unidade de seleção amostral é o domicílio, e as informações serão coletadas anualmente, por meio de questionários eletrônicos sobre todas as pessoas residentes no domicílio.

Entre os dois critérios de estratificação propostos, daremos maior ênfase, no planejamento da amostra, ao critério das Mesorregiões, com a subdivisão da Mesorregião de Triângulo/Alto Paranaíba nas Regiões de Planejamento que a compõem: Triângulo Mineiro e Alto Paranaíba, e da Mesorregião Metropolitana entre Belo Horizonte e o restante. Isto resulta em 14 estratos, que estão listados nas Tabelas 1 e 2.

Tabela 1. Total de domicílios e de setores censitários, tamanho médio de setor censitário MS (em domicílios), desvio-padrão, média e coeficiente de variação da renda do responsável por domicílio, por estrato (situação urbana).

ESTRATO	POPULAÇÃO	SETORES		RENDA		
		TOTAL	TAMA- NHO	DESVIO_P	MÉDIA	CV
1) Belo Horizonte	628.442	2.549	247	3.096,10	1.224,46	2,5285
2) Resto da Metropolitana	776.695	3.030	256	1535,42	536,54	2,8617
3) Central Mineira	88.006	304	289	1148,61	495,39	2,3186
4) Oeste de Minas	198.551	788	252	1495,19	585,42	2,5540
5) Campo das Vertentes	113.117	463	244	1244,06	595,18	2,0902
6) Zona da Mata	438.971	1.731	254	1352,22	635,02	2,1294
7) Sul/Sudoeste de Minas	484.299	1.910	254	2317,53	694,91	3,3350
8) Triângulo	339.663	1.381	246	2862,06	772,55	3,7046
9) Alto Paranaíba	139.251	616	226	4.006,27	652,50	6,1394
10) Noroeste de Minas	64.788	248	261	1634,76	518,49	3,1534
11) Norte de Minas	229.229	890	258	1.452,24	424,53	3,4202
12) Jequitinhonha	93.393	360	259	927,88	339,77	2,7309
13) Vale do Mucuri	62.925	261	241	1226,93	458,82	2,6741
14) Vale do Rio Doce	309.863	1.288	241	1381,66	561,29	2,4616
Minas Gerais	3.967.193	16.051	247	2220	683,15	3,2496

Tabela 2. Total de domicílios e de setores censitários, tamanho médio de setor censitário (em domicílios), desvio-padrão, média e coeficiente de variação da renda do responsável por domicílio, por estrato (situação rural).

ESTRATO	POPULAÇÃO	SETORES		RENDA		
		TOTAL	TAMANHO	DESVIO_P	MÉDIA	CV
1) Belo Horizonte	-	-	-	-	-	-
2) Resto da Metropolitana	84.830	759	112	945,17	310,07	3,0483
3) Central Mineira	16.448	161	102	4065,79	461,30	8,8137
4) Oeste de Minas	35.872	338	106	1605,64	366,15	4,3852
5) Campo das Vertentes	24.487	195	126	1048,52	312,63	3,3539
6) Zona da Mata	117.935	874	135	1820,38	320,79	5,6747
7) Sul/Sudoeste de Minas	135.322	984	138	1877,03	410,02	4,5779
8) Triângulo	31.952	297	108	1402,73	525,37	2,6700
9) Alto Paranaíba	27.865	267	104	1825,55	535,46	3,4093
10) Noroeste de Minas	22.015	245	90	3753,30	422,88	8,8755
11) Norte de Minas	115.187	811	142	378,86	160,00	2,3679
12) Jequitinhonha	64.384	470	137	388,33	157,44	2,4665
13) Vale do Mucuri	31.673	260	122	572,57	210,28	2,7229
14) Vale do Rio Doce	89.788	762	118	890,42	264,37	3,3681
Minas Gerais	797.758	6423	124		310,84	

Ao se garantir uma precisão minimamente aceitável em cada um dos estratos considerados, a agregação de dois ou mais deles para compor uma Região de Planejamento ou uma Mesorregião, a tendência é melhorar a estimativa. Este seria o caso das seguintes Regiões de Planejamento:

- a) Jequitinhonha/Mucuri, composta da Mesorregião de Mucuri e 88% da Mesorregião Jequitinhonha;
- b) Centro-Oeste, composta da Mesorregião Oeste de Minas e de 40% da Mesorregião Central Mineira;
- c) Sul, composta da Mesorregião Sul/Sudoeste de Minas e 25% da Mesorregião Campo das Vertentes, e

d) Central, composta da Mesorregião Metropolitana de BH, de 75% da Mesorregião Campo das Vertentes, de 60% da Mesorregião Central Mineira, e de 12% da Mesorregião Jequitinhonha.

O mesmo ocorreria tanto com a Mesorregião Triângulo/Alto Paranaíba, que teria sua estimativa melhorada ao se agregarem as duas Regiões de Planejamento que a compõem, quanto com a Região Metropolitana de Belo Horizonte, composta pela cidade de Belo Horizonte acrescida de alguns municípios da Mesorregião Metropolitana.

Tabela 3. Tamanho da população, por situação, em Domicílios (e Setores Censitários) nas Regiões de Planejamento, Mesorregiões e suas interseções, onde houve agregação ou desagregação.

MESO-REGIÃO	REGIÃO DE PLANEJAMENTO						
	RESTO DA CENTRAL	SUL DE MG	CO DE MG	ALTO PARA NAÍBA	TRIÂNGULO	JEQUITINHONHA/MUCURI	TOTAL
Resto da Metropolitana	U	258.408 (978)					
	R	56.098 (491)					
Campo das Vertentes	U	81.807 (333)	31310 (130)				113.117 (463)
	R	19.241 (150)	5246 (45)				24.487 (195)
Oeste de Minas	U			198.551 (788)			
	R			35.872 (338)			
Central Mineira	U	50.726 (178)		37.280 (126)			88.006 (304)
	R	11.723 (117)		4.725 (44)			16.448 (161)
Jequitinhonha	U	13.812 (52)				79.581 (308)	93.393 (360)
	R	4940 (44)				59.444 (426)	64.384 (470)
Vale do Mucuri	U					62.925 (261)	
	R					31.673 (260)	
Sul/Sudoeste De Minas	U		484.299 (1.910)				
	R		135.322 (984)				
Triângulo/Alto Paranaíba	U			139.251 (616)	339.663 (1.381)		478.914 (1.997)
	R			27.865 (267)	31.952 (297)		59.817 (564)
Total	U	404.753 (1.541)	515.609 (2.040)	235.831 (914)		142.506 (569)	
	R	92.002 (802)	140.568 (1.029)	40.597 (382)		91.117 (686)	

2.2 Alocação da amostra nos estratos

Com a finalidade de garantir uma precisão minimamente aceitável (erro relativo ou coeficiente de variação da média estimada, $CV \leq 0,15$) em cada estrato, será adotada alocação potência (Bankier, 1988), que tende a homogeneizar os erros relativos ou percentuais entre os estratos. Para os 14 estratos, $h = 1, \dots, 14$, a amostra total de $n = 18.000$ domicílios será distribuída de acordo com a regra:

$$n_h = n \frac{S_h Y_h^q / \bar{Y}_h}{\sum_h S_h Y_h^q / \bar{Y}_h} \quad (1), \text{ onde } S \text{ se refere ao desvio-padrão populacional, e } Y \text{ à renda}$$

total, ambos dos chefes de família, e quando se encontra com uma barra superior, trata-se da renda média. O parâmetro q é uma potência, que pode variar entre 0 e 1, e será determinada empiricamente. Podemos notar no gráfico da Figura 1 que, à medida que o valor de q aumenta, as estimativas agregadas para o Grande Norte pioram, ocorrendo o contrário para as estimativas agregadas para Minas Gerais.; além disso, quando olhamos os erros estimados para cada mesorregião (Figura 2), verificamos que a potência q não deve ultrapassar o valor 0,5, para que o CV se mantenha abaixo de 0,15, e que quanto menor a potência, maior a homogeneidade dos erros de estimação. Ficamos com a potência $q = 0,2$, pois os erros relativos do Grande Norte se mantêm nos níveis mínimos até este valor da potência, com algum ganho nos erros relativos de Minas Gerais.

Figura 1: Erro relativo da média estimada em função da potência q , para Grande Norte (GN) e Minas Gerais (MG).

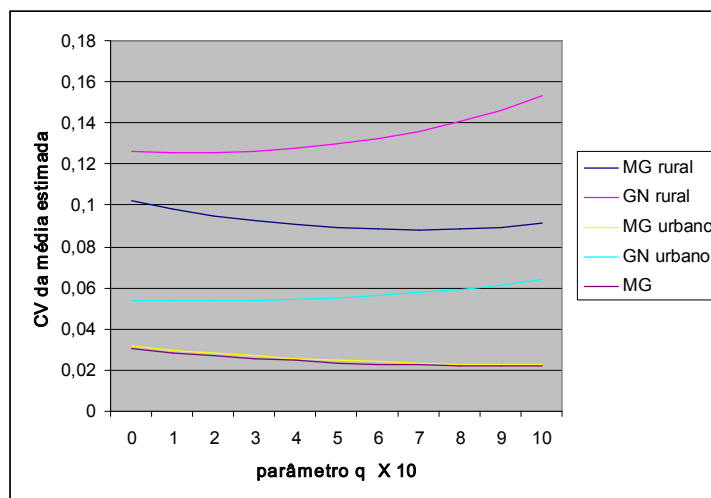
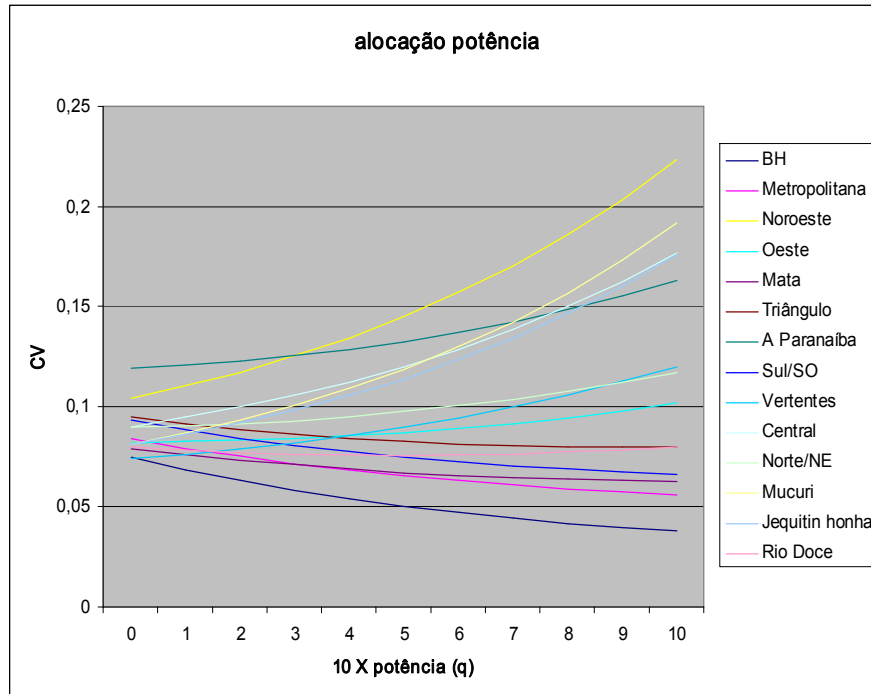


Tabela 4. Erro relativo da média estimada para o Grande Norte e Minas Gerais, nas situações urbana e rural.

	RURAL	URBANO
--	-------	--------

MG	9,50%	2,79%
Grande Norte	12,57%	5,35%

Figura 2. Erro relativo da média estimada em função da potência q , para cada mesorregião.



Cada estrato está subdividido em microrregiões, e cada microrregião, em situação urbana ou rural: consideramos uma sub-estratificação baseada na situação, que será utilizada para a alocação de Neyman da amostra calculada para o estrato, com o objetivo de minimizar a variância do erro de estimação em cada estrato. Podemos notar nas Tabelas 1 e 2, que existe uma diferença na renda média entre as situações urbana e rural. No caso de Belo Horizonte, foram considerados dois sub-estratos, definidos por um lado, através do distrito de Belo Horizonte (mais os sub-distritos da Pampulha do distrito de Venda Nova) e por outro lado, através dos demais distritos: Venda Nova e Barreiro.

Para cada situação s , do estrato h , a amostra de tamanho n_h será distribuída de acordo com a fórmula:

$$n_{hs} = n_h \frac{W_{hs} S_{hs}}{\sum_s W_{hs} S_{hs}} \quad (2), \text{ sendo } W \text{ o tamanho relativo do sub-estrato, e } S \text{ o desvio-padrão}$$

da renda.

A variância do erro de estimação de uma média no estrato h , utilizando a fórmula (2), fica:

$$\text{Var}(\bar{y}_h) = \frac{\left[\sum_s W_{hs} S_{hs} \right]^2}{n_h} \quad (3), \text{ cujo paralelo em Amostragem Aleatória Simples, sem}$$

levar em conta a correção para a população finita, é $\frac{S_h^2}{n_h}$. Assim, foi utilizado, na

fórmula (1), o valor entre colchetes da fórmula (3), no lugar do desvio-padrão S da renda em cada estrato h , uma vez que na fórmula (1) existe a premissa de utilização de AAS em cada estrato.

Tabela 4. Desvio-padrão médio ponderado por estrato, alocação percentual e tamanho da amostra nos sub-estratos, e tamanho da amostra e erro de estimação nos estratos.

ESTRATOS	$\sum_s W_{hs} S_{hs}$	n_{hs} / n_h		n_{hs}		Total n_h	Erro de estimação
Belo Horizonte	2757,61	VN/Barreiro 12,16%	BH 87,84%	VN/Barreiro 156	BH 1123	1279	6,30%
		rural	urbano	rural	urbano		
Metropolitana	1477,30	6,30%	93,70%	92	1369	1461	7,51%
Central	1607,97	39,82%	60,18%	432	652	1084	9,97%
Oeste	1512,09	16,25%	83,75%	177	913	1090	8,30%
Vertentes	1209,27	15,43%	84,57%	122	669	790	7,89%
Mata	1451,36	26,56%	73,44%	323	892	1214	7,33%
Sul/SO	2221,33	18,45%	81,55%	322	1421	1743	8,41%
Triângulo	2736,58	4,41%	95,59%	74	1615	1689	8,86%
A Paranaíba	3642,66	8,36%	91,64%	184	2015	2198	12,27%
Noroeste	2172,06	43,83%	56,17%	614	787	1402	11,74%
Norte/NE	1093,26	11,59%	88,41%	147	1119	1265	9,15%
Jequitinhonha	707,70	22,39%	77,61%	190	657	846	9,17%
Mucuri	1007,84	19,02%	80,98%	157	667	824	9,35%
Rio Doce	1271,30	15,74%	84,26%	175	938	1113	7,71%
			total	3007	14993	18000	MG 2,68%