

Pesquisa Distrital por Amostra de Domicílios Rural – PDAD Rural

NOTA METODOLÓGICA AMOSTRAL

A Pesquisa Distrital por Amostra de Domicílios Rural – PDAD Rural, realizada no Distrito Federal, pretende ser um instrumento de oferta de informações sobre o perfil socioeconômico e demográfico dos moradores e dos domicílios da área rural do território, com intuito de subsidiar o planejamento nas ações e tomadas de decisões de políticas públicas governamentais para área rurais.

1. Amostra

A base de endereços foi constituída utilizando-se o cadastro de consumidores da área rural de energia elétrica (CEB), água (CAESB), censo agropecuário (IBGE) e DF Legal¹. Com base em imagens de satélite e conhecimento técnico da equipe da EMATER, foram localizadas regiões com características rurais, não incluídas nos cadastros anteriores, e a equipe da Codeplan procedeu a listagem dessas áreas², utilizando georreferenciamento e visitas aos locais, adicionando o total de 496 endereços. Os endereços finais utilizados para a construção da amostra contemplam unidades 36.381 rurais em fevereiro de 2022.

Para o cálculo do tamanho da amostra foi considerado o método de populações finitas, isso ocorre quando a amostra tem um (n) maior ou igual a 5% do tamanho da população original. A relação a seguir corresponde à determinação do tamanho da amostra (n) com base na estimativa da proporção populacional:

$$n_i = \frac{N_i \hat{p} \hat{q} (Z_{\alpha/2})^2}{(N_i - 1) E^2 + \hat{p} \hat{q} (Z_{\alpha/2})^2},$$

¹ Os cadastros foram compartilhados pela Companhia Energética de Brasília – CEB, Companhia de Saneamento Ambiental do Distrito Federal – CAESB, Censo Agropecuário do IBGE de 2017 e Secretária de Estado de Proteção da Ordem Urbanística do Distrito Federal - DF LEGAL.

² As áreas correspondem ao Córrego do Urubu, Córrego da Onça, Núcleo Rural do Jerivá, Assentamento 8 de Março, Contagem e outras comunidades.

N_i corresponde ao total de endereços do estrato i ($i=1, \dots, 21$), já n_i a amostra correspondente ao estrato i ($i=1, \dots, 21$). O erro considerado foi de 5% com grau de confiança de 95%. A relação da fração de proporção é dada por:

$$F_i = \frac{n_i}{N_i},$$

em que n_i é o tamanho da amostra do estrato i e N_i domicílios na área rural do estrato i . Foi considerada uma margem de perda em torno de 16% da amostra, assim o resultado do tamanho amostral para cada estrato foi $n_i = n_i \cdot 1,16$.

A seleção da amostra, aplicada aos 21 estratos estabelecidos, seguiu o esquema aleatório, considerando a proporção de endereços dentro dos setores censitários por estrato, com o intuito de garantir amostras mais diversificadas e esparsas.

Tabela 1: Total de endereços com características rurais (N_i), tamanho da amostra (n_i), tamanho da amostra considerando perda estimada de 16%, cálculo da amostra final considerando a proporção de endereços por setor censitário segundo o estrato e porcentagem da amostra final sobre a amostra considerando a perda.

Estratos	N_i	n_i	$n_i \cdot 1,16$	Amostra Final	%
Agricultura Empresarial - Leste	3077	342	396	393	0,99
Agricultura Empresarial - Norte	1285	296	343	335	0,98
Agricultura Empresarial - Oeste	555	227	264	261	0,99
Agricultura Empresarial - Sul	1761	316	366	358	0,98
Agricultura Familiar - Central Adjacente I	256	154	178	177	0,99
Agricultura Familiar - Central Adjacente II	736	253	293	290	0,99
Agricultura Familiar - Leste	576	231	268	263	0,98
Agricultura Familiar - Norte	3198	343	398	378	0,95
Agricultura Familiar - Oeste	1750	315	366	352	0,96
Agricultura Familiar - Sul	285	164	190	187	0,98
Assentamento/ Agrovila - Leste	840	264	306	303	0,99
Assentamento/ Agrovila - Oeste	157	112	130	128	0,98
Assentamento/ Agrovila - Sul	100	80	92	90	0,98
Assentamento/Agrovila - Central Adjacente I	74	62	72	72	1,00
Assentamento/Agrovila - Norte	272	159	185	184	0,99
Características Urbanas - Central Adjacente I	89	72	84	81	0,96
Características Urbanas - Central Adjacente II	5240	358	415	409	0,99
Características Urbanas - Leste	8288	367	426	413	0,97
Características Urbanas - Norte	3443	346	401	385	0,96
Características Urbanas - Oeste	2833	338	393	386	0,98
Características Urbanas - Sul	1566	309	358	348	0,97
Total DF rural	36381	5108	5924	5793	0,98

Para o aprimoramento logístico das equipes que aplicaram o questionário, foi constituída uma amostra piloto, de tamanho máximo 100, da respectiva base de dados da amostra já estabelecida, sendo considerada apenas a Unidade de Planejamento Territorial Sul (UPT - Sul). Do total de endereços da amostra da região UPT – Sul, foi calculada a proporção do tamanho da amostra para o respectivo estrato e multiplicado por 100.

A seleção da amostra piloto empregou a mesma técnica já citada e utilizou os setores censitários para melhor distribuição espacial de acordo com o estrato.

Tabela 2: Total de endereços com características rurais da amostra na UPT Sul, proporção de domicílios e amostra piloto.

Estrato	n_i	Proporção	Amostra
Agricultura Empresarial - Sul	358	36	34
Agricultura Familiar - Sul	187	19	18
Assentamento/ Agrovila - Sul	90	9	7
Características Urbanas - Sul	348	35	37
Total Piloto rural	983	99	96

Tabela 3: Total de endereços com características rurais excluída amostra piloto.

Estrato	Total
Agricultura Empresarial - Leste	393
Agricultura Empresarial - Norte	335
Agricultura Empresarial - Oeste	261
Agricultura Empresarial - Sul	324
Agricultura Familiar - Central Adjacente I	177
Agricultura Familiar - Central Adjacente II	290
Agricultura Familiar - Leste	263
Agricultura Familiar - Norte	378
Agricultura Familiar - Oeste	352
Agricultura Familiar - Sul	169
Assentamento/ Agrovila - Leste	303
Assentamento/ Agrovila - Oeste	128
Assentamento/ Agrovila - Sul	83
Assentamento/Agrovila - Central Adjacente I	72
Assentamento/Agrovila - Norte	184
Características Urbanas - Central Adjacente I	81
Características Urbanas - Central Adjacente II	409
Características Urbanas - Leste	413
Características Urbanas - Norte	385
Características Urbanas - Oeste	386
Características Urbanas - Sul	311
Total	5697

Tabela 4: Total de endereços com características rurais, após o agrupamento dos estratos.

Estrato	Piloto	Amostra	Total
Agricultura Empresarial	34	1313	1347
Agricultura Familiar	18	1629	1647
Assentamento/Agrovilla	7	770	777
Características Urbanas	37	1985	2022
Total	96	5697	5793

Foram elaboradas duas listas contendo endereços de substituição que seriam acionadas em situações específicas para evitar o comprometimento da amostra. Essas listas consistiam em domicílios que estavam presentes na base de endereços final, porém não estavam presentes na base da amostra.

2. Processo de seleção da amostra de substituição

Os procedimentos para a seleção da amostra de substituição descritos destinam-se a garantir o não viés da seleção, assim além de serem considerados estratos que tornam as unidades mais homogêneas, considerando características intrínsecas da população, também podem ser considerados procedimentos que visam garantir a melhor distribuição da amostra dentro de cada estrato.

A seleção das amostras substitutivas foi estabelecida a partir da amostra de campo, em que foram ordenados os estratos e criado um id de identificação 1:5793. Os endereços presentes na amostra de campo foram retirados da base de endereços final com 36.381 observações, o que garantia uma nova base para a seleção da amostra de substituição com 30.588 observações, denominada endereços sem amostra. Foram calculados os máximos, mínimos e o total de observações por estrato dentro da amostra de campo, bem como dos endereços sem amostra, visando verificar se dentro de cada estrato poderia ser obtido uma amostra substitutiva de tamanho equivalente a amostra de campo por estrato.

Nota-se que para alguns estratos isso não seria possível devido a amostra de campo já contemplar praticamente todo o estrato, ficando apenas poucos endereços para serem utilizados de forma substitutiva. Para esses estratos que não tinham os endereços suficientes do tamanho amostral, foi feita uma nova base, contemplando os tamanhos

máximos, aqueles estratos que tinham um número de endereços que garantia selecionar uma amostra de substituição do tamanho da amostra de campo, bem como tamanhos mínimos. Nesse caso, foi substituído o tamanho amostral para o total de observações presentes na base endereços sem amostra.

O tamanho dos estratos da amostra de substituição continuaria com o tamanho do estrato da amostra de campo caso existissem endereços suficientes para criar duas amostras substitutivas naquele estrato, ou contemplaria o mínimo do total de endereços naquele estrato dividido por 2, caso o número de endereços restantes fosse menor que o da amostra de campo e conseqüentemente fosse insuficiente para criar duas amostras substitutivas. As amostras substitutivas utilizaram o mesmo desenho da amostra de campo, sendo utilizados 21 estratos e levando em consideração a proporcionalidade dos setores censitários pelo tamanho da amostra dentro de cada estrato.

Como mencionado, a primeira amostra substitutiva foi formada a partir da base endereços sem amostra, já a segunda amostra substitutiva foi constituída pelo resultado da base endereços sem amostra de campo e sem a substitutiva 1 apresentando um total de 25.902 observações. Essa base de endereços apresentava valores insuficientes de observações para 11 estratos, como opção foi escolhido retirá-los da base e formar uma amostra estratificada proporcional por setores apenas considerando os demais estratos (10 estratos), pois total de endereços seria suficiente para o desenho amostral e, por conseguinte reunir os endereços dos 11 estratos na amostra final.

A necessidade da criação de um identificador novo que associasse a amostra de campo a amostra substitutiva fica claro do ponto de vista estatístico, pois há necessidade de se considerar que aquele endereço que necessita ser substituído dentro do estrato considerado corresponde a um novo endereço na amostra substitutiva dentro do mesmo estrato considerado, evitando um viés de seleção amostral. Reduz-se, assim, o risco do pesquisador, por exemplo, substituir todos os endereços não localizados por endereços em áreas próximas, intencionando um ganho de tempo ou redução de custos no transporte.