

NOTA METODOLÓGICA

Prévia da População dos Municípios com base nos dados do Censo Demográfico de 2022 coletados até o dia 25/12/2022

Sumário

1	Introdução	2
2	Composição das populações municipais	2
2.1	Municípios coletados	2
2.2	Municípios não totalmente coletados	2
3	Definição de municípios e setores censitários coletados	3
4	Método de Correção da Lista Prévia	4
5	Tratamento dos dados.....	5
5.1	Pré-Crítica.....	5
5.2	Tratamento de não-resposta (domicílios ocupados sem entrevista)	5
5.2.1	Municípios coletados.....	5
5.2.2	Municípios não totalmente coletados	7
6	ANEXO.....	8
6.1	Categorias das variáveis para estratificação – método de Correção da Lista Prévia.....	8
6.2	Categorias das variáveis do modelo logístico - Reponderação.....	8

1 Introdução

Frente aos atrasos ocorridos no Censo Demográfico de 2022, não foi possível finalizar a coleta em todos os municípios do país a tempo de se fazer essa divulgação prévia dos resultados da pesquisa. Assim, para viabilizar o cumprimento das obrigações legais do IBGE em relação à divulgação das populações municipais do corrente ano, foram adotadas estratégias que visaram utilizar o máximo de informações coletadas na pesquisa, complementado, quando necessário, com estimativas realizadas com base nos próprios dados do Censo. Nas próximas seções serão apresentados os detalhes das metodologias utilizadas para viabilizar essa divulgação.

2 Composição das populações municipais

Para fins de composição das populações municipais, foi necessário separar os municípios em dois grandes grupos segundo os critérios que serão definidos na seção 3: municípios coletados e municípios não coletados.

2.1 Municípios coletados

Para os municípios considerados como coletados (4.410 municípios), a população considerada foi aquela observada na pesquisa com o devido tratamento da não-resposta, que foi realizado por meio de uma reponderação cuja metodologia será abordada na seção 5.2.1.

2.2 Municípios não totalmente coletados

Nos demais municípios não totalmente coletados (1.160 municípios), a população total foi composta pela soma da população observada nos setores considerados como coletados com a população estimada dos setores não totalmente coletados. A população oriunda dos setores coletados recebeu um tratamento de não resposta diferenciado que será descrito na seção 5.5.2. Já a parcela oriunda dos setores não totalmente coletados foi estimada por um método denominado de Método de Correção da Lista Prévia, descrito na seção 4. Uma vez estimada a população para um setor por meio dessa metodologia, essa estimativa foi comparada com os dados coletados até o dia 25/12/2022 no mesmo setor. Quando a estimativa estivesse abaixo do total populacional coletado, a população considerada foi aquela coletada até o momento, com o respectivo tratamento de não resposta (seção 5.2.2).

3 Definição de municípios e setores censitários coletados

O Setor Censitário é a menor porção de área utilizadas pelo IBGE para planejar, coletar e disseminar os resultados dos Censos e Pesquisas Estatísticas. Para o Censo Demográfico de 2022, o Brasil foi dividido em mais de 452 mil setores censitários, e cada um deles precisa passar por uma sequência de status até que o trabalho nesta unidade possa ser considerado como finalizado. Ainda assim, sempre que há a necessidade de alguma correção, os setores que poderiam ser considerados como finalizados podem retornar para o status de andamento para que os devidos ajustes sejam feitos.

No caso dos municípios, a lógica é semelhante. Um município só pode ser considerado como finalizado quando todos os seus setores já tiverem sido trabalhados e haja a indicação no sistema de que o mesmo já foi verificado e encerrado pela Superintendência Estadual. Depois disso, um comitê nacional avalia os dados desse município e pode solicitar correções, se necessário, em alguns casos, que haja reabertura de algum setor censitário para verificação.

Apesar de alguns estados e municípios estarem com a coleta bastante avançada, apenas algumas dezenas de municípios passaram por todas as etapas de verificação e poderiam, de fato, ser considerados como finalizados. Assim, para viabilizar a divulgação desses resultados prévios e fazer a composição das populações municipais de todos os municípios, foi necessário estabelecer alguns critérios para assumir que a coleta nos municípios estava finalizada ou não, pois isso tem efeito direto na metodologia utilizada para o tratamento da não resposta.

Como a avaliação do encerramento da coleta baseada nos status poderia não representar o real estágio da coleta, tanto em setores quanto nos municípios como um todo, optou-se por utilizar um critério baseado na lista prévia para definir quando se consideraria que a coleta estava ou não finalizada. A lista prévia nada mais é do que a relação de endereços e suas respectivas espécies (Exemplo: domicílio particular permanente ocupado, domicílio vago, estabelecimento de saúde etc.) que é utilizada como insumo nos Dispositivos Móveis de Coleta (DMC) dos recenseadores. Estas listas são baseadas nos dados observados no Censo Demográfico 2010 (cerca de 75%), atualizados durante a década nos setores onde foram realizadas a PNAD contínua. Em adição, foi realizado um grande esforço de melhoria, como etapa preparatória para a coleta do Censo Demográfico, nas áreas onde ocorreram grandes transformações no período. Estas áreas, alvo prioritário deste trabalho, foram identificadas principalmente através do cruzamento da base de dados do IBGE com a lista de endereços residenciais fornecida ao IBGE pela Agência Nacional de Energia Elétrica.

Assim, optou-se por definir que um município poderia ser considerado como coletado quando pelo menos 99% de sua lista prévia tivesse sido trabalhada e o percentual de domicílios ocupados sem entrevista fosse no máximo 15%.

No caso dos setores, adotou-se uma lógica semelhante, porém o percentual da lista prévia adotado para considerar o setor como coletado foi de pelo menos 87% e o percentual de domicílios ocupados sem entrevista fosse no máximo 27%.

Além disso, no caso dos setores censitários outras métricas indicativas da qualidade da coleta no setor foram utilizadas para que se considerasse o setor como coletado, a saber:

a razão entre o total de domicílios particulares permanentes ocupados atuais e os de da lista prévia (deve estar entre 0,75 e 2,5); razão entre a soma dos domicílios particulares permanentes vagos e de uso ocasionais atuais e os da lista prévia (deve ser menor ou igual a 2,5); razão entre a quantidade de endereços atuais e os da lista prévia (deve estar entre 0,8 e 2,5); razão entre a quantidade de endereços excluídos e a quantidade de endereços atuais (deve ser menor que 2,5); razão entre a quantidade de endereços incluídos e a quantidade de endereços atuais (deve ser menor ou igual a 0,7); e quantidade de reaberturas do setor (deve ser menor ou igual a 3).

4 Método de Correção da Lista Prévia

De forma bastante resumida, este método utiliza taxas observadas nos setores coletados no Censo para calcular fatores de correção para a lista prévia e estimar os totais de domicílios e pessoas nos setores não coletados.

O número de pessoas de um setor pode ser estimado com base no número de domicílios particulares permanentes ocupados (DPPO), no número de domicílios particulares improvisados ocupados (DPIO), no número de questionários em domicílios coletivos (QDC) e na média de moradores para cada um deles:

$$P_s = DPPO_s^* \times MMdppo_s^* + DPIO_s^* \times MMdpio_s^* + QDC_s^* \times MMqdc_s^* \quad (1)$$

Onde P_s é a população estimada do setor s ; $DPPO_s^*$ é o total estimado de DPPO no setor s ; $MMdppo_s^*$ é a média de moradores por DPPO estimada para o setor s ; $DPIO_s^*$ é o total estimado de DPIO no setor s ; $MMdpio_s^*$ é a média de moradores por DPIO estimada para o setor s ; QDC_s^* é o total estimado de QDC no setor s ; e $MMqdc_s^*$ é a média de moradores por DPPO estimada para o setor s .

Os totais de DPPO, DPIO e QDC de um setor podem ser estimados multiplicando-se o total observado na lista prévia por uma taxa (ou fator de correção) extraída de setores coletados com características semelhantes, ou seja, pertencentes ao mesmo estrato. As características utilizadas na formação dos estratos para fins da aplicação deste método foram a situação, o tipo e a característica do setor (ver classificações no Anexo). Os fatores de correção considerados foram calculados a partir da média dos valores observados desses fatores para os setores pertencentes ao mesmo estrato. O cálculo dos fatores de correção foi realizado respeitando-se também alguns estratos geográficos, começando nos subdistritos e indo até o total do Brasil, passando pelos distritos, municípios, Unidades da Federação e Regiões. Para um fator de correção ser considerado para fins de correção da lista prévia, foi estabelecido um número mínimo de 4 setores dentro do mesmo estrato. Sempre que este número mínimo não fosse atingido, o fator de correção considerado foi aquele calculado para o estrato geográfico imediatamente superior, desde que também tivesse no mínimo 4 setores.

Para estimar a média de moradores por domicílios e por questionário (para os domicílios coletivos), foram utilizados também os dados coletados de setores com características semelhantes, também estratificados de acordo com a situação, o tipo e a característica do setor e com o mesmo critério de no mínimo 4 setores no estrato.

5 Tratamento dos dados

5.1 Pré-Critica

Com o objetivo de garantir a consistência dos dados, foi realizada uma etapa de validação dos registros, que buscou identificar e tratar eventuais inconsistências. Os principais aspectos verificados neste procedimento foram:

- Questionários com pelo menos a lista de moradores preenchida corretamente;
- Questionários com registros tanto no banco de dados de domicílios quanto no de pessoas;
- Verificação de domicílios e/ou pessoas duplicadas nos bancos de dados;
- Verificação da convergência dos dados oriundas da Amostra presentes no Universo com os armazenados nos bancos específicos da Amostra;
- Comparação entre os registros de endereços e espécies, com os de questionários, com o intuito de verificar se os questionários estão associados a endereços confirmados.

5.2 Tratamento de não-resposta (domicílios ocupados sem entrevista)

Durante a operação de coleta do Censo Demográfico 2022, não foi possível realizar a entrevista em algumas unidades domiciliares ocupadas, mesmo após algumas visitas (não foi possível encontrar o morador ou houve uma recusa, por exemplo). Portanto, para se obter a população total dos municípios é necessário realizar algum ajuste.

5.2.1 Municípios coletados

Na divulgação dos resultados das populações municipais baseadas em estimativas e/ou dados do Censo Demográfico 2022, o método utilizado para tratar a não resposta nos municípios considerados finalizados, a fim de obtermos o total populacional, foi o da ponderação das unidades domiciliares respondentes. Desta maneira, o total populacional de um município m será dado por:

$$P_m = \sum_{i=1}^{N_m^*} w_{im} \times y_{im} \quad (2)$$

Onde w_{im} é o peso e y_{im} é o total de moradores da unidade domiciliar i com entrevista realizada do município m . N_m^* é o número de domicílios ocupados com entrevista do município m .

Para se obter w_{im} foi utilizado o método estratificado (ou pós-estratificado) da modelagem da propensão de resposta. Nesse método, proposto por Little (1986), os fatores de ajuste são calculados em pós-estratos criados a partir das propensões de resposta estimadas. Um modelo de regressão logística foi utilizado para prever essa propensão de resposta. Esse método parte do princípio que estão disponíveis covariáveis, tanto para as unidades respondentes quanto para as não respondentes, que possam estar correlacionadas com a probabilidade de um domicílio responder o Censo Demográfico.

Desta maneira, definindo $r_i = 1$ se o domicílio i respondeu o Censo e $r_i = 0$ se o domicílio i não respondeu e seja $P(r_i = 1) = \theta_i$ a probabilidade de um domicílio i do município m responder. Podemos escrever o modelo logístico da seguinte forma:

$$\text{Logit}(P(r_i = 1)|X_k) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (2)$$

$$\text{Ln}\left(\frac{\theta_i}{1 - \theta_i}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (3)$$

Onde $X_i, i = 1 \dots k$ são as covariáveis utilizadas: faixas de tamanho populacional do município, tipo de espécie do domicílio, situação do domicílio, tipo de setor, característica do setor, número de mudanças de espécie/ situação de entrevista do domicílio e uma indicadora de adensamento do setor¹.

Foi utilizado um modelo para cada Unidade da Federação e as probabilidades estimadas foram então classificadas em 10 grupos da seguinte maneira: menos de 70%, de 70% a menos de 75%, de 75% a menos de 80%, de 80% a menos de 85%, de 85% a menos de 90%, de 90% a menos de 94%, de 94% a menos de 97%, de 97% a menos de 98%, de 98% a menos de 99% e 99% ou mais.

Cada município foi dividido em 4 categorias, a saber:

1. Domicílios particulares permanentes não pertencentes a Povos de Comunidades Tradicionais;
2. Domicílios particulares improvisados não pertencentes a Povos de Comunidades Tradicionais;
3. Terras Indígenas ou setor de aldeia indígena; e
4. Territórios Quilombolas ou agrupamentos quilombolas.

Em cada uma dessas categorias foram utilizadas as 10 classes de probabilidades construídas a partir do modelo logístico para a definição dos pós-estratos. Caso não houvesse entrevista realizada em algum pós-estrato, esses foram colapsados, e assim foram criados os pós-estratos para o ajuste de não resposta.

Dentro de cada pós-estrato g , o fator de ajuste (peso) de cada domicílio ocupado com entrevista foi calculado da seguinte maneira:

$$w_{mg} = \frac{dt_{mg}}{dr_{mg}} \quad (4)$$

¹ As categorias utilizadas em cada variável estão descritas no anexo 1.

Onde dt_{mg} é o total de domicílios ocupados do pós-estrato g do município m e dr_{mg} é o total de domicílios ocupados do pós-estrato g do município m que responderam o Censo Demográfico.

O ajuste para os domicílios ocupados coletivos foi realizado separadamente. Foram criados pós-estratos diretamente em cada município, considerando apenas o tipo de espécie de cada domicílio (asilo, hotel, pensão, alojamento, penitenciária, etc.). A fórmula utilizada é a mesma observada em (4).

5.2.2 Municípios não totalmente coletados

O tratamento da não resposta para os municípios não totalmente coletados se deu seguindo uma lógica parecida com a do Censo Demográfico de 2010. Para os setores coletados, caso o percentual de domicílios ocupados sem entrevista (DPPOSE) fosse superior a 5%, os dados foram corrigidos por meio da média de moradores observadas nos setores pertencentes ao mesmo estrato (situação, tipo e característica do setor, além da estratificação geográfica). Foram adotados os mesmos critérios utilizados na estimação da média de moradores para a aplicação do Método de Correção da Lista prévia, ou seja, adotou-se como critério o número mínimo de 4 setores no estrato, adotando a média do estrato geográfico superior caso esse mínimo não fosse atingido.

Caso o percentual de domicílios ocupados sem entrevista (DPPOSE) fosse menor ou igual a 5%, foi aplicado um fator de correção à média estimada para os setores com mais de 5% de DPPOSE. Esses fatores de correção foram construídos no nível de Unidades da Federação de forma estratificada (situação, tipo e característica do setor). O fator de correção foi dado pela razão entre a média de moradores dos domicílios que em algum momento foram classificados como DPPOSE e posteriormente foi realizada uma entrevista, dividido pela média geral de moradores por DPPO.

6 ANEXO

6.1 Categorias das variáveis para estratificação – método de Correção da Lista Prévia

Situação do domicílio

- 01-Área Urbanizada
- 02-Área Não Urbanizada
- 03-Área Urbana Isolada
- 04-Rural-Aglom Rural De Ext Urb
- 05-Rural-Aglom Rural Isol Povoado
- 06-Rural-Aglom Rural Isol Núcleo
- 07-Rural-Aglom Rural Isol Outros Aglom
- 08-Rural-Zona Rural Excl Aglom Rural
- 09-Massas d'água

Tipo de setor

- 00-Comum Ou Não Especial
- 01-Espec De Aglom Subnormal(Favelas E Similares)
- 02-Espec De Quartéis, Bases Militares, Etc.
- 03-Espec De Alojам, Acamp, Etc.
- 04-Espec De Embarc, Barcos, Navios, Etc.
- 05-Espec De Aldeia Indígena
- 06-Espec De Penitenc, Colôn Penais, Presíd, Cadeias, Etc.
- 07-Espec De Asilos, Orfan, Conv, Hosp, Etc.
- 08-Espec De Assent Rural
- 09-Agrupamentos Quilombolas

Característica do Setor

- 1 – Normal
- 2 – Indígena
- 3 – Quilombola

6.2 Categorias das variáveis do modelo logístico - Reponderação

Faixas de tamanho populacional do município

- 1 – Menos que 13.584 habitantes
- 2 – De 13.585 habitantes até 23.772 habitantes
- 3 – De 23.773 habitantes até 37.356 habitantes
- 4 – De 37.357 habitantes até 71.316 habitantes
- 5 – De 71.317 habitantes até 115.464 habitantes

6 – De 115.465 habitantes até 156.216 habitantes

7 – 156.217 habitantes ou mais

Tipo de espécie do domicílio

1 – Casa

2 – Casa de vila ou em condomínio

3 – Apartamento

4 – Habitação em casa de cômodos ou cortiço

5 – Outro tipo de domicílio particular permanente

6 – Tenda ou barraca de lona

7 – Outro tipo de domicílio particular improvisado

Situação do domicílio

1 – Urbano

2 – Rural

Tipo de setor

1 – Comum ou não especial

2 – Aglomerado subnormal

3 – Aldeia Indígena

4 – Agrupamentos quilombolas

5 – Outro tipo de setor

Característica do Setor

1 – Normal

2 – Indígena

3 – Quilombola

Número de mudanças de espécie/situação de entrevista do domicílio

1 – Uma

2 – Duas

3 – Três

4 – Quatro

5 – Cinco ou mais

Referência:

Little, R.J.A. (1986). Survey Nonresponse Adjustments for Estimates of Means. International Statistical Review, 54, 139–157.