**Descrição do plano de amostragem**

**Apresentação**

Este documento consiste na descrição do plano de amostragem referente ao levantamento quantitativo da pesquisa Building the Barricades, realizado entre julho e dezembro de 2019, no conjunto de favelas da Maré, no Rio de Janeiro, Brasil.

Segundo o Censo Populacional da Maré (Redes da Maré, 2019), as 16 favelas da Maré abrigavam 139.073 moradores no ano de 2013, entre os quais, 97.878 com 18 ou mais anos de idade. Na presente pesquisa, foram entrevistadas 1.211 pessoas, selecionadas aleatoriamente entre os moradores maiores de 18 anos de idade.

**População de pesquisa**

A população de pesquisa corresponde ao conjunto de moradores em domicílios das 16 favelas do conjunto de favelas da Maré maiores de 18 anos de idade. O cadastro de endereços do Censo Populacional da Maré utilizado para a seleção da amostra possui 47.776 endereços.

**Estratificação geográfica da população de pesquisa**

A população de pesquisa foi estratificada em três estratos geográficos, compostos por agrupamentos de favelas da Maré, delimitados segundo a localização, as características da habitação e as dinâmicas sociais das favelas, além de vias de acesso e instalações públicas comuns.

Os estratos geográficos 1, 2 e 3 abrangem, respectivamente, quatro, nove e três favelas da Maré, conforme apresentado no Quadro 1, a seguir.

| Quadro 1 – Favelas da Maré abrangidas em cada estrato geográfico | |
| --- | --- |
| Estratos geográficos | Favelas que constituem cada estrato |
| Área 1 | Nova Holanda, Parque Maré, Parque Rubens Vaz e Parque União |
| Área 2 | Baixa do Sapateiro, Conjunto Bento Ribeiro Dantas, Conjunto Esperança,  Conjunto Pinheiros, Morro do Timbau, Nova Maré, Salsa e Merengue, Vila do João, Vila dos Pinheiros |
| Área 3 | Marcílio Dias, Parque Roquete Pinto, Praia de Ramos |

A distribuição da população de pesquisa por faixas de idade e sexo em cada estrato geográfico da amostra é apresentada no Quadro 2, a seguir.

| Quadro 2 – Distribuição da população maior de 18 anos de idade por faixas de idade, segundo os estratos geográficos e sexo | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| Estratos geográficos e sexo | Total | Faixas de idade | | | | |
| 18 a 29 anos | 30 a 49 anos | 50 a 65 anos | 66 ou mais anos | Idade ignorada |
| Área 1 | 38.313 | 12.749 | 17.048 | 5.845 | 2.287 | 384 |
| Masculino | 18.648 | 6.270 | 8.517 | 2.786 | 891 | 184 |
| Feminino | 19.588 | 6.472 | 8.521 | 3.057 | 1.395 | 143 |
| Ignorado | 77 | 7 | 10 | 2 | 1 | 57 |
| Área 2 | 47.415 | 14.744 | 21.412 | 7.890 | 2.795 | 574 |
| Masculino | 22.843 | 7.211 | 10.578 | 3.660 | 1.138 | 256 |
| Feminino | 24.470 | 7.518 | 10.825 | 4.221 | 1.657 | 249 |
| Ignorado | 102 | 15 | 9 | 9 | – | 69 |
| Área 3 | 12.150 | 3.433 | 5.743 | 2.154 | 746 | 74 |
| Masculino | 5.833 | 1.683 | 2.803 | 1.032 | 288 | 27 |
| Feminino | 6.295 | 1.746 | 2.939 | 1.122 | 458 | 30 |
| Ignorado | 22 | 4 | 1 | – | – | 17 |
| Fonte: REDES DA MARÉ. Censo Populacional (2013). Rio de Janeiro: Redes da Maré, 2019. | | | | | | |

**Seleção da amostra nos estágios do plano de amostragem**

O plano de amostragem empregado na pesquisa é estratificado e selecionado em dois estágios: domicílios particulares e morador adulto (18 anos ou mais de idade).

No primeiro estágio, os domicílios foram selecionados por amostragem inversa (Haldane, 1945; Vasconcellos *et al*., 2005; Vasconcellos *et al*., 2013).

No segundo estágio, um morador adulto foi selecionado com equiprobabilidade entre os moradores adultos do domicílio selecionado.

A amostragem inversa é um procedimento sequencial de amostragem, largamente empregado pelos sistemas que realizam contagens sanguíneas, e tem como vantagens o fato de eliminar, ou pelo menos mitigar bastante, as não respostas observadas nas pesquisas domiciliares clássicas.

Na amostragem domiciliar clássica, é determinado um número fixo de domicílios a entrevistar por estrato ou unidade amostral, que são selecionados por algum método de seleção aleatória. Então, os domicílios selecionados são visitados e as respostas ou não respostas (recusas, domicílios vagos, de uso ocasional, demolidos etc.) observadas; as não respostas ocasionam uma perda na amostra, perda essa diferente em cada estrato ou unidade amostral. Para lidar com este problema, três alternativas podem ser empregadas:

(1) substituição dos domicílios com não resposta que, independentemente da regra de substituição definida, sempre gera vieses de seleção;

(2) aumento do tamanho de amostra pela aplicação de uma taxa de não resposta esperada, o que não é o mais adequado porque a taxa de não resposta é diferenciada por estrato e ocasiona desperdícios, quando a amostra acaba maior do que o previsto, ou não resolve o problema, quando a taxa de não resposta é maior do que a esperada; ou

(3) uso de *oversampling*, que é difícil de aplicar e costuma aumentar o custo da coleta, visto que a compensação de perda geralmente é feita depois de terminado o trabalho de coleta, obrigando o retorno ao campo para complementar a amostra.

A amostragem inversa altera esse paradigma mudando o tamanho da amostra para um número fixo de sucessos (entrevistas completas realizadas), independentemente do número de domicílios que são visitados para a obtenção da entrevista. Sua forma de randomização é feita na definição de uma ordem aleatória de endereços a visitar.

Assim, na amostragem inversa, o entrevistador vai recebendo a relação de endereços e os visita sequencialmente até atingir o número de previstas ou esgotar a área (estrato ou unidade amostral) de pesquisa.

O método foi originalmente descrito por Haldane (1945) para realização de contagens de elementos do sangue e transposto para a amostragem domiciliar por Vasconcellos e colegas em 2005.

**Tamanho da amostra**

Considerando que não há informações na população da pesquisa sobre a prevalência das principais variáveis estudadas, o plano de amostragem partiu de outra estratégia. Foi escolhida a proporção mínima de 3% (Pmin = 0,03), para a qual o erro relativo da estimativa deveria ser de, no máximo, 60% (dR = 60%), o que corresponde a uma margem de erro relativa de 1,8%, com um nível de confiança de 95% (1-α= 0,95).

De acordo com Cochran (1977) e supondo uma amostragem aleatória simples sem reposição (AAS), o tamanho da amostra necessário para estimar proporções iguais ou superiores a Pmin com erro relativo não superior a dR no nível de confiança 1-α é dado por:

(1)

Com os parâmetros fixados, o uso da expressão em (1), conduz a um tamanho de amostra de 346 adultos (ou domicílios, visto que, na presente pesquisa, só um adulto foi selecionado no segundo estágio da amostra) em cada estrato geográfico.

Como cada estrato geográfico constitui um domínio de estimação controlada, foi mantido o mesmo tamanho de amostra para cada um, assegurando em cada estrato a estimação de proporções de 3% com erro relativo de 60% no nível de confiança de 95%.

No entanto, como um conjunto de perguntas do questionário era referido a pessoas de até 65 anos de idade e não foi considerado operacional fixar uma idade máxima para o entrevistado, o tamanho de amostra foi aumentado para 400 adultos em cada estrato geográfico, conduzindo a um tamanho final de amostra de 1.200 domicílios (ou adultos), somadas as amostras dos três estratos.

Esse pequeno aumento do tamanho da amostra conduz a uma pequena redução da margem de erro. Cabe ressaltar, também, que os resultados para o conjunto da Maré apresentam margem de erro menor do que para cada estrato em particular. No entanto, o relevante é determinar os parâmetros de cálculo da amostra, pois, a rigor, os erros amostrais devem ser estimados para cada variável a partir dos dados que foram coletados.

**Esquema probabilístico da amostra**

A probabilidade de inclusão na amostra do domicílio i do estrato geográfico h, representada por , é dada por:

, onde, (2)

th é o total de domicílios do estrato h;

vh é o total de domicílios visitados do estrato h;

ah é o total de domicílios com adultos (domicílios elegíveis) do estrato h; e

eh é o total de domicílios entrevistados do estrato h.

Note que as três razões acima indicam, respectivamente, a probabilidade de o domicílio ser visitado, a probabilidade de o domicílio ser elegível e a probabilidade de o domicílio ser entrevistado. O “- 1” na terceira fração corresponde à regra de parada do procedimento sequencial, ou seja, a perda de um grau de liberdade porque o último domicílio elegível é também o último entrevistado.

A probabilidade de o adulto Ahij ser incluído na amostra, dado que o domicílio dhi foi selecionado, representada por, é dada por:

, onde, (3)

nahi é o número de adultos do domicílio hi do estrato h.

Assim, a probabilidade de inclusão de um adulto qualquer , representada por , é dada pela expressão abaixo:

(4)

O peso natural do desenho corresponde ao inverso da expressão em (4), observando-se que a segunda razão deve ser eliminada para evitar expandir o número de domicílios elegíveis para a população total. Assim, o peso natural do desenho de um adulto qualquer, representado por é dado em (5)

(5)

**Calibração dos pesos amostrais**

Nas pesquisas domiciliares, a amostra representa a população de domicílios e isso sempre provoca vieses na estimação da população por sexo e grupos de idade, em decorrência da diversidade do número de moradores dos domicílios e da estrutura etária e por sexo dos mesmos.

Além disso, todas as probabilidades de inclusão da amostra referem-se a cadastros defasados no tempo. Nessa pesquisa, as probabilidades de inclusão refletem a população da Maré em 2013, portanto seis anos antes do momento da pesquisa.

Assim, a literatura internacional recomenda calibrar os pesos amostrais de forma que possam ser ajustados a estimativas independentes da população por sexo e idade, cuja data de referência seja próxima ao período da pesquisa (Silva, 2004).

No caso, foi usado um estimador de pós-estratificação para calibrar os pesos amostrais, de modo que o peso calibrado do adulto j do domicílio i do estrato h representado por , é igual ao seu peso do desenho multiplicado por um fator de calibração, definido para o pós-estrato p, representado por . Assim, o peso calibrado é dado por:

(6)

, onde, (7)

é o valor da estimativa independente do número de pessoas no pós-estrato p; e

é o total estimado do número de pessoas no pós-estrato p.

Os oito pós-estratos são definidos pelo cruzamento do sexo com a classe de idade (18 a 29 anos; 30 a 49 anos; 50 a 65 anos; 66 anos ou mais). Os dados da estimativa independente da população da Maré é apresentada no Quadro 3.

Como os dados censitários disponíveis mais recentes são de 2013 e a amostra foi coletada em 2019, houve a estimação do tamanho da população de cada estrato, por sexo e idade, em 1º de julho de 2019.

Partindo das estimativas de população[[1]](#footnote-1) por Unidade da Federação e idade simples, produzidas pelo IBGE para os anos de 2010 (Censo Demográfico 2010, do IBGE), 2013 (ano do Censo da Maré) e 2019 (momento da pesquisa), foi utilizado o método da tendência linear (Madeira e Simões, 1972) para estimar a população da Maré para 1º de julho de 2019, apresentada no Quadro 3.

| Quadro 3 – Estimativa da população maior de 18 anos de idade da Maré para 1º de julho de 2019 por sexo, segundo faixas de idade | | | |
| --- | --- | --- | --- |
| Classe de idade | Total | Homens | Mulheres |
| Total | 101.549 | 49.435 | 52.114 |
| 18 a 29 anos | 30.603 | 15.186 | 15.417 |
| 30 a 49 anos | 45.911 | 22.911 | 23.000 |
| 50 a 65 anos | 17.750 | 8.418 | 9.332 |
| 66 anos ou mais | 7.285 | 2.920 | 4.365 |

**Procedimentos operacionais e dados da amostra efetiva**

Dois procedimentos operacionais devem ser pormenorizados nesta seção: (1) procedimento de randomização da ordem de visita aos domicílios; e (2) procedimento de seleção do adulto a ser entrevistado no domicílio.

O procedimento de randomização dos endereços foi feito colocando, a começar, todos os endereços de cada estrato na ordem alfabética por favela, logradouro, número no logradouro e complemento. Para cada endereço, foi gerado um número aleatório uniforme no intervalo (0;1). Em seguida, os endereços de cada estrato foram ordenados na ordem crescente desses números aleatório gerados. De uma forma mais científica, foi gerada uma permutação aleatória dos endereços de cada estrato.

Cada entrevistador recebeu um conjunto pequeno de endereços a visitar e, ao completar o trabalho nestes, recebeu um novo lote de endereços. Com esse procedimento é possível assegurar que os entrevistadores respeitem a ordem aleatória de percurso.

O Quadro 4 indica o número de endereços da Maré, quantos endereços foram passados aos entrevistadores e quantos foram efetivamente entrevistados.

| Quadro 4 – Número de endereços disponíveis no Censo da Maré, número de endereços fornecidos aos entrevistadores e número de domicílios efetivamente entrevistados | | | |
| --- | --- | --- | --- |
| Estratos geográficos | Total de endereços | Endereços fornecidos aos entrevistadores | Domicílios entrevistado |
| Total | 47.776 | 3.136 | 1.211 |
| Área 1 | 19.231 | 1.130 | 406 |
| Área 2 | 22.336 | 1.054 | 400 |
| Área 3 | 6.209 | 952 | 405 |

Observa-se no Quadro 4, que todos os endereços de imóveis entraram no procedimento de amostragem, ou seja, tiveram probabilidade não nula de serem selecionados. De fato, não foi feita qualquer separação prévia dos endereços em função de sua destinação em 2013, tendo em vista que seu uso poderia ter sido alterado entre 2013 e 2019, quando a pesquisa foi realizada.

O procedimento de seleção do adulto a ser entrevistado no domicílio selecionado foi baseado na seleção de amostras aleatórias simples para cada possível tamanho de domicílio. Se o domicílio tinha um único morador adulto, este foi selecionado. Se tinha dois adultos, foi selecionada, a partir dos números 1 e 2, uma amostra de tamanho 1, ou seja, ora foi selecionado o número 1 ora o número 2. Se o domicílio tinha três moradores, foi selecionada uma amostra de tamanho 1 a partir dos números 1, 2, 3. Este procedimento foi repetido até 15 moradores. Assim, foi elaborada uma tabela de seleção com 48.000 linhas e 16 colunas, conforme exemplo no Quadro 5.

Para a seleção do morador do domicílio, o primeiro passo é ordenar por idade os moradores acima de 18 anos (adultos), do mais velho para o mais novo. Após esse registro, a linha correspondente ao domicílio é consultada. Se o entrevistador estiver no primeiro domicílio, a consulta é à primeira linha; se o segundo domicílio, na segunda linha e, assim, sucessivamente. Na linha consultada, é observada a coluna correspondente ao número de moradores adultos do domicílio – se, por exemplo, cinco moradores adultos, a coluna 5. Assim, o cruzamento linha x coluna indica o número de ordem do morador selecionado. Se o número indicado for o número 2, o segundo mais velho é o morador selecionado.

| Quadro 5 – Quadro de seleção do adulto a ser entrevistado | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nº de ordem do domicílio selecionado | Número de moradores adultos do domicílio | | | | | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 1 | 1 | 1 | 2 | 1 | 2 | 3 | 4 | 2 | 3 | 8 | 11 | 8 | 8 | 10 | 9 |
| 2 | 1 | 1 | 1 | 3 | 3 | 5 | 7 | 2 | 1 | 8 | 1 | 3 | 5 | 4 | 15 |
| 3 | 1 | 2 | 1 | 2 | 1 | 2 | 7 | 2 | 2 | 3 | 4 | 3 | 2 | 3 | 5 |
| 4 | 1 | 2 | 1 | 3 | 3 | 5 | 1 | 5 | 2 | 3 | 10 | 10 | 6 | 6 | 10 |
| 5 | 1 | 2 | 3 | 2 | 5 | 4 | 4 | 7 | 5 | 2 | 4 | 12 | 11 | 3 | 2 |
| 6 | 1 | 1 | 2 | 4 | 4 | 5 | 2 | 5 | 5 | 9 | 2 | 12 | 10 | 6 | 2 |
| 7 | 1 | 1 | 2 | 4 | 3 | 2 | 6 | 6 | 8 | 10 | 1 | 9 | 9 | 6 | 2 |
| 8 | 1 | 2 | 3 | 1 | 4 | 3 | 6 | 1 | 6 | 2 | 3 | 6 | 4 | 11 | 1 |
| 9 | 1 | 1 | 2 | 3 | 2 | 6 | 1 | 2 | 2 | 3 | 3 | 10 | 10 | 1 | 14 |
| 10 | 1 | 1 | 2 | 1 | 4 | 1 | 3 | 5 | 2 | 9 | 6 | 3 | 6 | 10 | 11 |
| 11 | 1 | 1 | 3 | 2 | 2 | 1 | 3 | 8 | 8 | 3 | 8 | 10 | 3 | 4 | 2 |
| 12 | 1 | 1 | 2 | 2 | 5 | 1 | 5 | 5 | 6 | 1 | 3 | 2 | 1 | 9 | 5 |
| 13 | 1 | 2 | 3 | 1 | 4 | 3 | 1 | 1 | 2 | 3 | 7 | 4 | 13 | 2 | 13 |
| 14 | 1 | 1 | 2 | 3 | 2 | 3 | 1 | 8 | 4 | 10 | 8 | 4 | 6 | 8 | 1 |
| **. . .** | | | | | | | | | | | | | | | |

A busca por domicílios a serem entrevistados, ao longo do procedimento sequencial de amostragem inversa, mostrou os resultados apresentados no Quadro 6.

| Quadro 6 – Resultado das visitas aos domicílios | | | | |
| --- | --- | --- | --- | --- |
| Resultados das visitas aos domicílios | Total | Área 1 | Área 2 | Área 3 |
| Total | 3.136 | 1.130 | 1.054 | 952 |
| Entrevistas realizadas | 1.211 | 406 | 400 | 405 |
| Pessoa selecionada ausente em todas as visitas | 201 | 52 | 84 | 65 |
| Pessoa selecionada ausente (viagem, internação etc.) | 12 | 4 | 3 | 5 |
| Recusa da pessoa selecionada | 32 | 9 | 16 | 7 |
| Nenhum adulto em condições de ser entrevistado | 14 | 5 | 5 | 4 |
| Recusa do domicílio | 169 | 66 | 60 | 43 |
| Domicílio fechado ou inacessível | 1.048 | 414 | 330 | 304 |
| Domicílio não localizado ou inexistente (\*) | 352 | 130 | 124 | 98 |
| Domicílio vago (desocupado, viagem, obra etc.) | 72 | 33 | 29 | 10 |
| Unidade não residencial | 25 | 11 | 3 | 11 |

(\*) Esses casos decorrem da precariedade dos endereços em favelas, com muitos becos sem nome e domicílios sem numeração. A instrução de coleta, na dúvida de acessar o domicílio do endereço selecionado, era classificar o domicílio como inexistente ou não encontrado.

Observa-se que, somando os três estratos, foi necessário visitar 3.136 endereços para ter uma amostra de 1.211 domicílios entrevistados. Isso significa que 61% dos endereços visitados conduziram a não respostas, o que seria um problema grave se fosse usada a amostra domiciliar clássica.

Ao final da operação de coleta de dados nos 1.211 domicílios entrevistados, obteve-se uma amostra de 1.211 adultos, cuja distribuição por faixa etária é apresentada no Quadro 7.

| Quadro 7 – Número de adultos por classe de idade, segundo os estratos geográficos | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Estrato geográfico | Total | Faixas de idade e sexo | | | | | | | |
| 18 a 29 anos | | 30 a 49 anos | | 50 a 65 anos | | 66 anos ou mais | |
| Homens | Mulheres | Homens | Mulheres | Homens | Mulheres | Homens | Mulheres |
| Total | 1.211 | 113 | 168 | 190 | 303 | 116 | 166 | 49 | 106 |
| Área 1 | 406 | 32 | 66 | 68 | 101 | 36 | 53 | 14 | 36 |
| Área 2 | 400 | 38 | 50 | 63 | 107 | 34 | 63 | 18 | 27 |
| Área 3 | 405 | 43 | 52 | 59 | 95 | 46 | 50 | 17 | 43 |

O Quadro 7 indica os vieses de seleção comuns em pesquisas domiciliares. O total de mulheres entrevistadas foi de 743 (61,3%) e de adultos com 66 anos ou mais foi de 155 (12,8%). Por outro lado, o Quadro 3 indica que a população tem 51,3% de mulheres e 7,2% de adultos com 66 anos ou mais, o que comprova o viés de disponibilidade (a amostra tem mais mulheres e mais pessoas mais velhas porque são mais disponíveis para entrevista).

É exatamente esse viés de disponibilidade associado ao aumento da população total que justificam a calibração dos pesos amostrais.

**Referências**

COCHRAN, WG. Sampling techniques. 3rd Ed. New York: John Wiley & Sons, 1977.

HALDANE, JBS. On a method of estimating frequencies. Biometrika 1945; 33:222-225.

MADEIRA, J. L.; SIMÕES, C. C. da S. Estimativas preliminares da população urbana e rural, segundo as unidades da federação, de 1960/1980, por uma nova metodologia. Revista Brasileira de Estatística, 1972, 33(129): 3-11.

Redes da Maré. Censo Populacional da Maré (2013). Rio de Janeiro: Redes da Maré, 2019. ISBN: 978-85-61382-08-7. CDD 304.66. CDU 351.778.57. Disponível em: <https://www.redesdamare.org.br/media/downloads/arquivos/CensoMare_WEB_04MAI.pdf>

SILVA, PLN. Calibration Estimation: When and Why, How Much and How. Rio de Janeiro: IBGE (Textos para discussão da Diretoria de Pesquisa, nº 15).

VASCONCELLOS, MTL; SILVA, PLN; SZWARCWALD CL. Sampling design for the World Health Survey in Brazil. Cadernos de Saúde Pública 2005; 21(S):S89-S99.

VASCONCELLOS MTL; SILVA PLN; ANJOS LA. Sample design for the Nutrition, Physical Activity and Health Survey (PNAFS), Niterói, Rio de Janeiro, Brazil. Estadística 2013; 65(184):47-61.

1. disponíveis em [ftp://ftp.ibge.gov.br/Projecao\_da\_Populacao/Projecao\_da\_Populacao\_2018 /projecoes\_2018\_populacao\_idade\_simples\_2010\_2060.xls](ftp://ftp.ibge.gov.br/Projecao_da_Populacao/Projecao_da_Populacao_2018/projecoes_2018_populacao_idade_simples_2010_2060.xls)) [↑](#footnote-ref-1)