

Mercado de Trabalho

conjuntura e análise

ANO 28 | abril de 2022

73

ipea

Governo Federal

Ministério da Economia

Ministro Paulo Guedes

ipea Instituto de Pesquisa
Econômica Aplicada

Fundação pública vinculada ao Ministério da Economia, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiros – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente

Erik Alencar de Figueiredo

Diretor de Desenvolvimento Institucional

André Sampaio Zuvanov

**Diretor de Estudos e Políticas do Estado,
das Instituições e da Democracia**

Flavio Lyrio Carneiro

**Diretor de Estudos e Políticas
Macroeconômicas**

Marco Antônio Freitas de Hollanda Cavalcanti

**Diretor de Estudos e Políticas Regionais,
Urbanas e Ambientais**

Nilo Luiz Saccaro Júnior

**Diretor de Estudos e Políticas Setoriais de Inovação
e Infraestrutura**

João Maria de Oliveira

Diretor de Estudos e Políticas Sociais

Herton Ellery Araújo

**Diretor de Estudos e Relações Econômicas
e Políticas Internacionais (substituto)**

José Eduardo Malta de Sá Brandão

Assessor-chefe de Imprensa e Comunicação

André Reis Diniz

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

URL: <http://www.ipea.gov.br>

Mercado de Trabalho: conjuntura e análise

CORPO EDITORIAL

Editor Responsável

Sandro Pereira Silva

Membros

Carlos Henrique Leite Corseuil

Felipe Mendonça Russo

Lauro Ramos

Sandro Sacchet de Carvalho

Equipe de Apoio

Bruna de Souza Azevedo

Carolina Lopes de Carvalho Vital

Gabriela Carolina Rezende Padilha

Máira Albuquerque Penna Franca

Leandro Pereira da Rocha

© Instituto de Pesquisa Econômica Aplicada – ipea 2022

Mercado de trabalho : conjuntura e análise / Instituto de Pesquisa Econômica Aplicada; Ministério do Trabalho. – v.1, n.0, (mar.1996)- .- Brasília: Ipea: Ministério do Trabalho, 1996-

Irregular (de 1996-2008); Trimestral (de 2009-2012); Semestral (a partir de 2013).

Título da capa: Mercado de Trabalho: conjuntura e análise

ISSN 1676-0883

1. Mercado de Trabalho. 2. Estatísticas do Trabalho. 3. Brasil. 4. Periódicos. I. Instituto de Pesquisa Econômica Aplicada. II. Brasil. Ministério do Trabalho.

CDD 331.1205

DOI: <http://dx.doi.org/10.38116/bmt73>

As publicações do Ipea estão disponíveis para download gratuito nos formatos PDF (todas) e EPUB (livros e periódicos). Acesse: <http://www.ipea.gov.br/portal/publicacoes>

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério da Economia.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

SOBRE A MONTAGEM E A IDENTIFICAÇÃO DOS PAINÉIS DA PNAD CONTÍNUA¹

Rafael Guerreiro Osorio²

1 INTRODUÇÃO

A Pesquisa Nacional por Amostra de Domicílios (PNAD) Contínua nasceu da fusão da antiga PNAD Anual e da Pesquisa Mensal de Emprego (PME). A PNAD era realizada uma vez por ano, com representatividade quase nacional, e a partir de 2004; e a PME tinha periodicidade mensal, abrangendo apenas seis regiões metropolitanas. A PME era uma pesquisa de mercado de trabalho típica, visando à produção de indicadores frequentes sobre o mercado de trabalho; e a PNAD, um exemplo de pesquisa amostral de uso geral, investigando outros temas além do mercado de trabalho, como educação, migração, fecundidade, renda e outros não regulares, abordados em questionários suplementares.

A PME possuía uma amostra longitudinal: uma vez sorteado, o domicílio era visitado por quatro meses consecutivos, ficava sem receber visitas nos oito meses seguintes e depois era visitado mais quatro vezes. Curiosamente, o Instituto Brasileiro de Geografia e Estatística (IBGE) nunca divulgou os identificadores de grupos domésticos e indivíduos que permitiriam o uso do painel da PME. Ribas e Soares (2008) divulgaram um método para fazer a identificação com os dados disponíveis.

O esquema dos painéis da PNAD Contínua é diferente. Os domicílios são visitados cinco vezes consecutivas, trimestralmente. Dependendo do trimestre e do número da visita, são aplicados questionários suplementares que aumentam as possibilidades de uso dos dados. Porém, todas as distribuições oficiais da PNAD Contínua são voltadas para o uso dos microdados transversais. Embora os domicílios sejam visitados cinco vezes, as bases de dados, trimestrais ou anuais, só contam com uma visita ao domicílio. Todavia, o IBGE ainda não divulgou os identificadores que permitem a utilização dos dados em painel, o que dificulta a elaboração de estudos com análises longitudinais.

Esta dificuldade não tem impedido que pesquisadores montem, por conta própria, os painéis e elaborem suas análises. Os domicílios estão identificados, e, apesar de trabalhoso, é relativamente fácil montar os painéis de domicílios. Porém, os grupos domésticos e os seus indivíduos não estão identificados, e um domicílio pode ser ocupado por mais de um grupo doméstico; e, mesmo se o grupo doméstico for o mesmo em todas as visitas, seus membros

1. DOI: <http://dx.doi.org/10.38116/bmt73/nt3>

2. Pesquisador na Diretoria de Estudos e Políticas Sociais (Disoc) do Ipea. *E-mail*: <rafael.osorio@ipea.gov.br>.

podem mudar. Sem a marcação do IBGE, é preciso identificar os grupos domésticos e os indivíduos usando características como a data de nascimento e o sexo.

Contudo, a identificação pela data de nascimento e pelo sexo enfrenta alguns obstáculos. O principal são as datas de nascimento ignoradas. Além disso, pode haver inconsistências no registro dessas duas informações. Diferenças entre entrevistas dificultam a identificação. Nesse sentido, este estudo apresenta um método para realizar a identificação dos grupos domésticos e dos indivíduos que também leva em consideração o número de ordem e a condição no domicílio, o que aumenta a eficácia da identificação em relação a abordagens que se valham apenas da data de nascimento, da idade estimada e do sexo.

Todas as etapas da montagem e da identificação dos painéis descritas aqui são realizadas pelo aplicativo Pynad,³ que também se encarrega de fazer o *download* dos arquivos da PNAD Contínua e de mantê-los sincronizados, automatizando todo o processo. O aplicativo funciona em Windows e Linux, devendo ser instalado com o gerenciador de pacotes da linguagem de programação *python*.

2 OS MICRODADOS DA PNAD CONTÍNUA

Antes de usar os microdados da PNAD Contínua, é fundamental ler o documento que descreve sua amostragem (IBGE, 2014) e o manual de entrevista (IBGE, 2016) que explica a operação de campo, o objetivo das perguntas e o conteúdo das respostas. A amostra é complexa e começa com a seleção das unidades primárias de amostragem (*clusters*), distribuídas por centenas de estratos, definidos em um primeiro nível por áreas geográficas, e dentro das áreas pela homogeneidade estatística de características socioeconômicas. Em seguida, são selecionados, aleatoriamente, catorze domicílios em cada unidade primária de amostragem (UPA). Nos microdados, uma UPA pode ter menos de catorze domicílios, visto que as bases não têm registros para os domicílios que foram visitados, mas não pode ter menos entrevistados do que isso. Visitas sem entrevista não excluem um domicílio da amostra.

Continuamente, há sempre cinco painéis de domicílios em visitação, e as visitas de cada painel são uniformemente distribuídas no trimestre, com um intervalo de três meses entre as visitas. Em um trimestre, aproximadamente 20% dos domicílios são visitados pela primeira vez, e 20% estão na quinta visita. No trimestre subsequente, os domicílios visitados pela quinta vez no trimestre anterior saem da amostra e são substituídos por um novo painel de domicílios. A amostra de cada painel possui aproximadamente 3 mil UPAs e 42 mil domicílios. Por conta das visitas sem entrevista, nos microdados, a quantidade de domicílios por painel varia em torno de 40 mil a 41 mil.

O IBGE não divulga os microdados por painel, apenas em distribuições transversais, com uma entrevista de cada domicílio. Atualmente, há três tipos de distribuições oficiais da PNAD Contínua, conforme a seguir.

3. Disponível em: <<https://bit.ly/382R89F>>.

- 1) Trimestral:
 - a) questionário básico de todas as entrevistas do trimestre;
 - b) entrevistas com domicílios de cinco painéis (identificados pelo número da visita); e
 - c) projeção de população no meio do trimestre.
- 2) Anual de trimestre:
 - a) questionário básico e suplementar do trimestre (educação, tecnologias da informação e comunicação – TICs, turismo);
 - b) entrevistas com domicílios de cinco painéis (identificados pelo número da visita); e
 - c) projeção de população no meio do trimestre.
- 3) Anual de visita:
 - a) questionário básico e suplementar da primeira ou quinta visita em um ano;
 - b) entrevistas com domicílios de quatro painéis (identificados por ano e trimestre); e
 - c) projeção de população no meio do ano.

2.1 Separação dos painéis de domicílios

Como os microdados possuem quatro ou cinco painéis, o primeiro passo para usá-los é separá-los; e, o segundo, agregar as entrevistas de seus domicílios e as variáveis das entrevistas. Esta é uma tarefa trabalhosa, visto que os registros das entrevistas estão distribuídos por, no mínimo, cinco arquivos de microdados. Considere-se o caso do painel iniciado no segundo trimestre de 2016, que tem registros em dez arquivos de microdados:

- nos cinco arquivos trimestrais, do segundo trimestre de 2016 até o segundo trimestre de 2017;
- no arquivo anual de primeiras visitas em 2016 e no de quintas visitas em 2017; e
- nos arquivos anuais do segundo trimestre de 2016 e de 2017 (educação) e do quarto trimestre de 2016 (TICs).

Porém, há variáveis repetidas em diferentes arquivos. As bases trimestrais e as anuais de trimestre são redundantes: a base trimestral, a primeira a ser divulgada pelo IBGE com as variáveis do questionário básico, está contida na base anual de trimestre, que incorpora as variáveis do questionário suplementar. Os microdados anuais de educação de 2016 e 2017 contêm todas as variáveis da base trimestral do segundo trimestre; os de TICs, todas as do quarto trimestre. Assim, “somente” sete arquivos precisam ser lidos para obter os registros completos de pessoas nas cinco visitas do painel do segundo trimestre de 2016. Nesse sentido, as variáveis dos registros da:

- primeira visita vêm da base anual do segundo trimestre (educação) de 2016 e da base anual das primeiras visitas de 2016;
- segunda visita vêm da base trimestral do terceiro trimestre de 2016;

- terceira visita vêm da base anual do quarto trimestre (TICs) de 2016;
- quarta visita vêm da base trimestral do primeiro trimestre de 2017; e
- quinta visita vêm da base anual do segundo trimestre de 2017 e da base anual das quintas visitas de 2017.

Nas bases de microdados, cada domicílio é identificado por três variáveis: *upa*, *v1008* e *v1014*. O código da UPA a identifica na Amostra Mestra (IBGE, 2014) e vale para todas as pesquisas do Sistema Integrado de Pesquisas Domiciliares. A variável *v1008* (que marca os domicílios de 1 a 14) identifica o número do domicílio dentro da UPA, e a *v1014*, por sua vez, distingue os grupos de rotação da amostra. Esta é importante nos trimestres em que há dois grupos de rotação, visto que uma UPA pode ser selecionada mais de uma vez. Nas bases trimestrais, o painel é identificado a partir da variável *v1016*, que contém o número da visita; na base anual de visita, a identificação ocorre pela variável *trimestre*.

2.2 Identificação dos grupos domésticos e dos indivíduos

Os domicílios dos painéis da PNAD Contínua são construções usadas para moradia, casas e apartamentos. Nada garante que o grupo doméstico residente seja o mesmo nas cinco visitas. Um domicílio pode ter entrevistas com até cinco grupo domésticos diferentes. Os grupos domésticos, por sua vez, mudam de composição. Perdem membros quando casais se separam, filhos saem de casa, idosos falecem; e ganham membros por nascimento, uniões conjugais e outros eventos. É possível, portanto, um domicílio ter quatro pessoas entrevistadas em todas as visitas, mas serem indivíduos diferentes, de grupos domésticos distintos. Um indivíduo pode ser representado por até cinco registros de pessoas.

As entrevistas da PNAD Contínua são realizadas com questionários digitais e dispositivos eletrônicos de coleta, iniciando pela elaboração do quadro de moradores do domicílio na visita (IBGE, 2016). A princípio, define-se a pessoa responsável pelo grupo doméstico e pergunta-se o seu sexo, a sua data de nascimento e a cor ou raça. Para os demais moradores, os mesmos dados são registrados depois de estabelecida a categoria de relação com a pessoa responsável: cônjuge, filho etc. Incluindo a pessoa responsável, são dezenove categorias de “condição no domicílio”. É o entrevistado – ou os entrevistados – que deve definir qual morador é a pessoa responsável, e nada impede que, a cada visita, indivíduos diferentes sejam classificados nesta categoria.

O quadro de moradores do aplicativo de coleta registra também o nome e o sobrenome dos moradores, mas estes dados não são divulgados. Quando o quadro de moradores é finalizado, o aplicativo de coleta atribui um número de ordem para cada morador, em função da condição no domicílio e da data de nascimento. Nesse contexto, o quadro é atualizado em todas as entrevistas subsequentes, e as perguntas sobre a relação com a pessoa responsável, o sexo, a data de nascimento e a cor ou raça são feitas desconsiderando as respostas em entrevistas anteriores. Logo, no conjunto de entrevistas, um indivíduo pode ter registros de pessoa com variações no sexo, na data de nascimento, na cor ou raça, na condição no domicílio e no número de ordem.

Analicamente, há três fontes de variação dos dados que permitem identificar os indivíduos no conjunto de pessoas entrevistadas em todas as visitas a um domicílio: i) a mudança da pessoa responsável, que implica alterações na condição no domicílio e na ordem; ii) o acréscimo ou exclusão de moradores, que pode provocar alterações na condição no domicílio e na ordem atribuída pelo aplicativo da coleta; e iii) a retificação da declaração ou do registro de condição no domicílio, sexo ou data de nascimento.

O quadro 1 ilustra, para um grupo doméstico hipotético, como a escolha da pessoa responsável pode alterar substantivamente a condição no domicílio e o número de ordem dos indivíduos em diferentes entrevistas. A classificação dos demais moradores também pode variar, segundo a percepção das relações pelo grupo doméstico. Na primeira entrevista, Mário ou a pessoa entrevistada poderia considerar que Bruno, filho de sua enteada Mariana, é seu neto; ou, na quarta entrevista, dona Maria poderia ser classificada como avó de João.

QUADRO 1

Variação da ordem e condição no domicílio, segundo a seleção da pessoa responsável

| Morador | Data de nascimento | Número de ordem (v2003) e condição no domicílio (v2005) | | | | |
|---------------------|--------------------|---|---------------------|---------------------|---------------|---------------------|
| | | Entrevista 1 | Entrevista 2 | Entrevista 3 | Entrevista 4 | Entrevista 5 |
| Maria Ferreira | 15/2/1936 | 6 Sogra | 6 Mãe | 1 Responsável | 7 Convivente | 7 Avó |
| Mário Cruz | 20/11/1952 | 1 Responsável | 2 Cônjuge | 3 Genro | 2 Pai | 3 Padrasto |
| Teresa Ferreira | 4/5/1956 | 2 Cônjuge | 1 Responsável | 2 Filha responsável | 3 Madrasta | 4 Mãe |
| João Cruz | 11/9/1977 | 4 Filho responsável | 5 Enteado | 7 Convivente | 1 Responsável | 6 Irmão |
| Mariana Ferreira | 19/7/1982 | 5 Enteada | 4 Filho responsável | 4 Neta | 5 Irmã | 1 Responsável |
| Júlia Ferreira Cruz | 26/6/1992 | 3 Filha de ambos | 3 Filha de ambos | 5 Neta | 4 Irmã | 5 Irmã |
| Bruno Ferreira | 15/8/2012 | 7 Agregado | 7 Neto | 6 Bisneto | 6 Agregado | 2 Filho responsável |

Elaboração do autor.

Suponha-se que Mário faleceu depois da primeira entrevista. Ele seria excluído do quadro de moradores na segunda entrevista. Além de a exclusão alterar os números de ordem, sem haver cônjuge, Júlia passaria a ser classificada como filha da pessoa responsável.

Outra possibilidade seria Mário ser sempre apontado como a pessoa responsável, mas João e Mariana terem sido registrados erroneamente na primeira entrevista como filhos de Mário e Teresa. O erro pode ser de quem foi entrevistado – um erro de declaração – ou do entrevistador – um erro de registro. Se o erro fosse retificado em uma entrevista posterior, isto geraria alterações na condição de domicílio e no número de ordem.

O sexo e a cor ou raça podem mudar de uma entrevista para outra, por erro ou não. Há indivíduos que trocam de sexo, ou que mudam sua percepção sobre sua categoria racial. No caso da cor ou raça, há a complicação adicional introduzida pelo fato de poder ser outro morador, não necessariamente o indivíduo, que responde sobre a classificação racial.

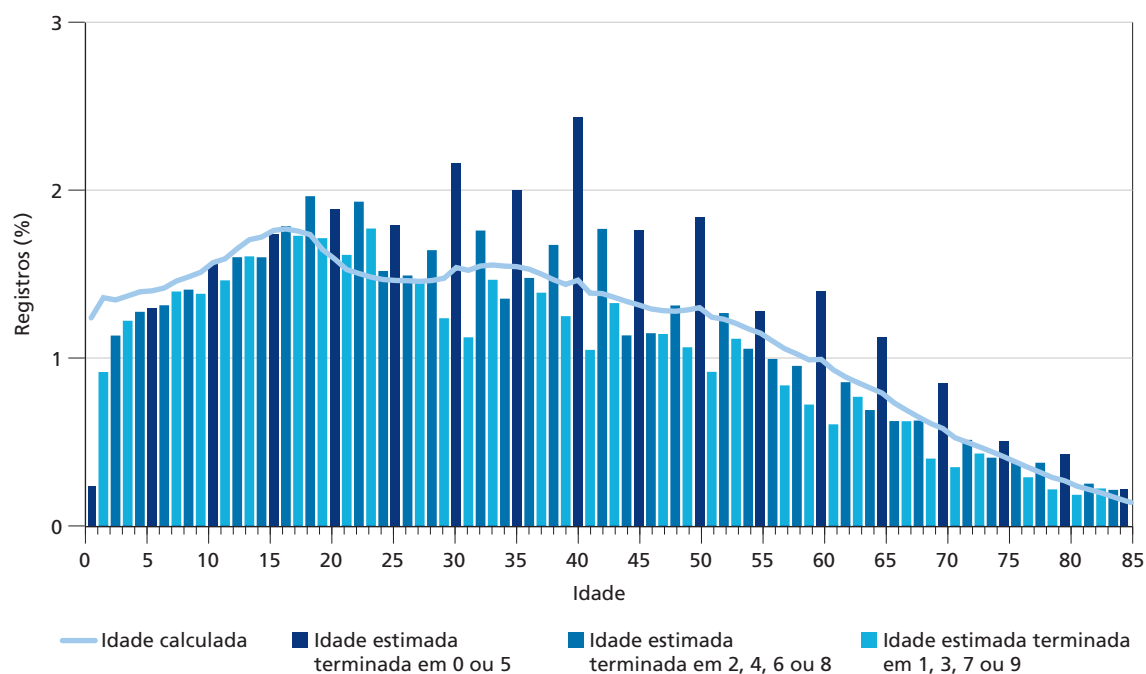
Um adolescente pode se considerar pardo, mas sua mãe enquadrá-lo como branco, ou o contrário. Efetivamente, na PNAD Contínua, a cor ou raça apresenta muita variação entre entrevistas, ao ponto de inviabilizar seu uso como critério de identificação.

Todavia, o maior obstáculo à identificação dos indivíduos é a data de nascimento. A data de nascimento não varia, mas é um dado muito suscetível a erros de declaração ou registro. O entrevistado pode dizer que um morador nasceu no dia 3 e o entrevistador entender como 6. É comum um morador saber o aniversário e a idade de outro, mas não o ano de nascimento, e haver erros no cálculo deste – pelo entrevistado ou entrevistador. De fato, é frequente a situação em que o dia e o mês não mudam, mas o ano de nascimento é trocado na entrevista do primeiro trimestre do ano subsequente ao do início de um painel.

A data de nascimento apresenta ainda o problema de poder ser ignorada e registrada como 99/99/9999. Quando a data de nascimento é ignorada, a idade em anos completos na data de referência da entrevista deve ser estimada pelo entrevistado, ou mesmo pelo entrevistador. Porém, a idade estimada é muito afetada pelo problema da “atração por dígito”, a tendência a arredondar a idade informada para um múltiplo de cinco ou um número par. Nesse contexto, o gráfico 1 compara a distribuição de idades calculadas a partir da data de nascimento com a de idades estimadas, nos quase 18 milhões de registros de pessoas acumulados de 2012 a 2020 pela PNAD Contínua.

GRÁFICO 1

Atração por dígito na estimação de idades para pessoas com data de nascimento ignorada



Fonte: Microdados trimestrais da PNAD Contínua (período 2012-2020). Disponível em: <<https://bit.ly/3ILD Afr>>. Elaboração do autor.

Para lidar com estes obstáculos, o processo de identificação é dividido em três etapas. A primeira envolve um tratamento das datas de nascimento ignoradas; a segunda, a identificação dos grupos domésticos; a terceira, a identificação dos indivíduos dos grupos domésticos. O processo de identificação parte dos painéis separados e é feito por domicílio, considerando o conjunto dos registros de pessoas entrevistadas em cada domicílio.

2.3 Tratamento das idades ignoradas

A primeira etapa para o tratamento das datas de nascimento ignoradas é a atribuição do ano de nascimento implicado pela idade estimada. O ano de nascimento imputado é a diferença entre a idade estimada e o ano inicial do painel, por exemplo, em um painel com primeira entrevista em 2016, ao registro de pessoa com idade estimada em 66 anos, imputa-se nascimento em 1950.

A segunda etapa é verificar se uma pessoa com data de nascimento ignorada em uma ou mais entrevistas teve sua data de nascimento declarada em outra entrevista, mediante os procedimentos a seguir.

- 1) Os registros de pessoas do domicílio são divididos em dois grupos, com data de nascimento ignorada e com data conhecida.
- 2) Para cada pessoa com data de nascimento ignorada, os procedimentos são os descritos adiante.
 - a) Filtra-se a lista das pessoas com data de nascimento conhecida para selecionar potenciais doadoras do atributo, observando os seguintes critérios:
 - não podem ser pessoas registradas na mesma entrevista;
 - devem ser do mesmo sexo;
 - o ano de nascimento da doadora pode ter até três anos de diferença absoluta para o ano de nascimento estimado; e
 - devem ter a mesma condição no domicílio, ou pertencer a alguns conjuntos de categorias de condição no domicílio.
 - b) Havendo doadoras, são ordenadas pela diferença entre o seu ano de nascimento e o ano de nascimento estimado.
 - c) A pessoa com data ignorada recebe a data de nascimento da doadora mais próxima (a primeira da lista); se não houver doadora, permanece com a data ignorada.

No terceiro critério do passo 2.a, a diferença de até três anos permite lidar com a atração por dígito (gráfico 1). Para um ano de nascimento qualquer, o ano de nascimento correspondente à idade múltipla de cinco mais próxima estará no intervalo com amplitude de sete anos centrado no ano: [ano - 3, ano + 3].

No quarto critério do passo 2.a, permite-se a variação da relação com a pessoa responsável. As variações permitidas abrangem as mudanças que ocorrem se o cônjuge vira pessoa responsável, ou o contrário, para as categorias mais frequentes de condição no domicílio:

- pessoas responsáveis e cônjuges, $v2005 \in \{1, 2, 3\}$;
- filhos e enteados, $v2005 \in \{4, 5, 6\}$; e
- pai, mãe, sogros, $v2005 \in \{8, 9\}$.

2.4 Identificação dos grupos domésticos

Os indivíduos de grupos domésticos distintos possuem conjuntos de entrevistas sem interseção. Se um domicílio tem indivíduos com registros de pessoa nas primeira e segunda visitas, e outros com registros nas terceira, quarta e quinta visitas, há dois grupos domésticos. Se um indivíduo tem registros de pessoa em todas as entrevistas, há apenas um grupo doméstico, não importando quão radicais possam ser as mudanças na sua composição.

A identificação do grupo doméstico depende da identificação dos indivíduos, mas prescinde da identificação de todos. A maior parte dos grupos domésticos tem ao menos um indivíduo com características constantes em todas as suas entrevistas. Sua presença facilita a identificação do grupo doméstico, a despeito de eventuais mudanças e erros de declaração ou registro nos atributos dos outros moradores. Como registros com erros poderiam levar a fusões ou separações impróprias de grupos domésticos, é preferível identificar os grupos domésticos antes dos indivíduos.

O domicílio com uma entrevista só pode ter um grupo doméstico. Para os domicílios com mais de uma entrevista, o algoritmo de identificação:

- separa as pessoas do domicílio por entrevista;
- seleciona a entrevista com maior número de pessoas com data de nascimento registrada;
- cria um grupo doméstico com as pessoas da entrevista selecionada;
- atribui ao grupo doméstico todas as pessoas das outras entrevistas em que havia ao menos uma pessoa com o sexo e a data de nascimento idênticos aos de uma pessoa já atribuída ao grupo doméstico; e
- havendo entrevistas não atribuídas, retorna ao passo 2 citado anteriormente, desconsiderando as atribuídas.

2.5 Identificação dos indivíduos

A identificação dos indivíduos é feita por grupo doméstico. Para cada grupo doméstico com mais de uma entrevista, verifica-se inicialmente se tem a mesma quantidade de registros de pessoas em todas as entrevistas. Para os grupos domésticos de tamanho constante, os registros de pessoas das diferentes entrevistas são agregados em indivíduos segundo o sexo, a data de

nascimento, a condição no domicílio e o número de ordem. Se os indivíduos têm pessoas em todas as entrevistas, sua quantidade é igual ao tamanho do grupo doméstico. Neste, que é o caso mais simples e de menor incerteza na identificação, em todas as entrevistas o grupo doméstico é composto por conjuntos de pessoas idênticas nessas quatro características. Os indivíduos desses grupos pertencem à classe 1 de identificação.

Se um grupo doméstico tem tamanho constante, mas há indivíduos que não estão presentes em todas as entrevistas, isto pode ser devido a: mudanças na condição no domicílio e no número de ordem provocadas pela troca da pessoa responsável; uma alteração na composição – sai um membro, entra outro; ou um erro no registro de sexo, data de nascimento ou condição no domicílio. Para esses grupos, os registros de pessoas das diferentes entrevistas são agregados em indivíduos, segundo o sexo e a data de nascimento. Se os indivíduos estão presentes em todas as entrevistas, o grupo doméstico é composto por conjuntos de pessoas idênticas nessas duas características, mas ao menos um tem variações na condição no domicílio e número de ordem. Os indivíduos desses grupos pertencem à classe 2 de identificação.

Grupos domésticos que mudam de tamanho ou composição, e/ou cujas pessoas possuam atributos com erros variados de declaração ou registro, precisam de maior atenção. Essas mudanças implicam indivíduos que não estão presentes em todas as entrevistas de seu grupo doméstico. O conjunto de regras, a seguir, é empregado para identificar os seus indivíduos:

- agregar em indivíduos as pessoas de entrevistas diferentes com o mesmo sexo e data de nascimento; e
- separar os indivíduos em dois grupos, o dos que estão em todas as entrevistas do grupo doméstico e o dos que não estão (indivíduos que aparecem em todas as entrevistas pertencem à classe 3).

Os indivíduos que aparecem em algumas entrevistas, classificados inicialmente como pertencentes à classe 4, podem ser o resultado de uma mudança de composição do grupo doméstico. Se há mais de um indivíduo classe 4 com conjuntos mutuamente exclusivos de entrevistas, podem ser fragmentos de um indivíduo. Quando isto ocorre, são reclassificados como classe 5. O indivíduo fragmentado não aparece em algumas entrevistas por conta de pessoas com erros ou mudanças de declaração ou registro de data de nascimento ou sexo. Um indivíduo pode estar fragmentado em até cinco indivíduos. Para verificar se os indivíduos que não aparecem em todas as entrevistas são fragmentos, realizam-se os procedimentos a seguir.

- 1) Elaborar a lista das combinações possíveis de indivíduos sem entrevistas em comum.
- 2) Comparar os atributos sexo, dia, mês e ano de nascimento, e condição no domicílio das pessoas dos indivíduos de cada combinação e registrar as diferenças entre as pessoas de todos os pares de indivíduos.

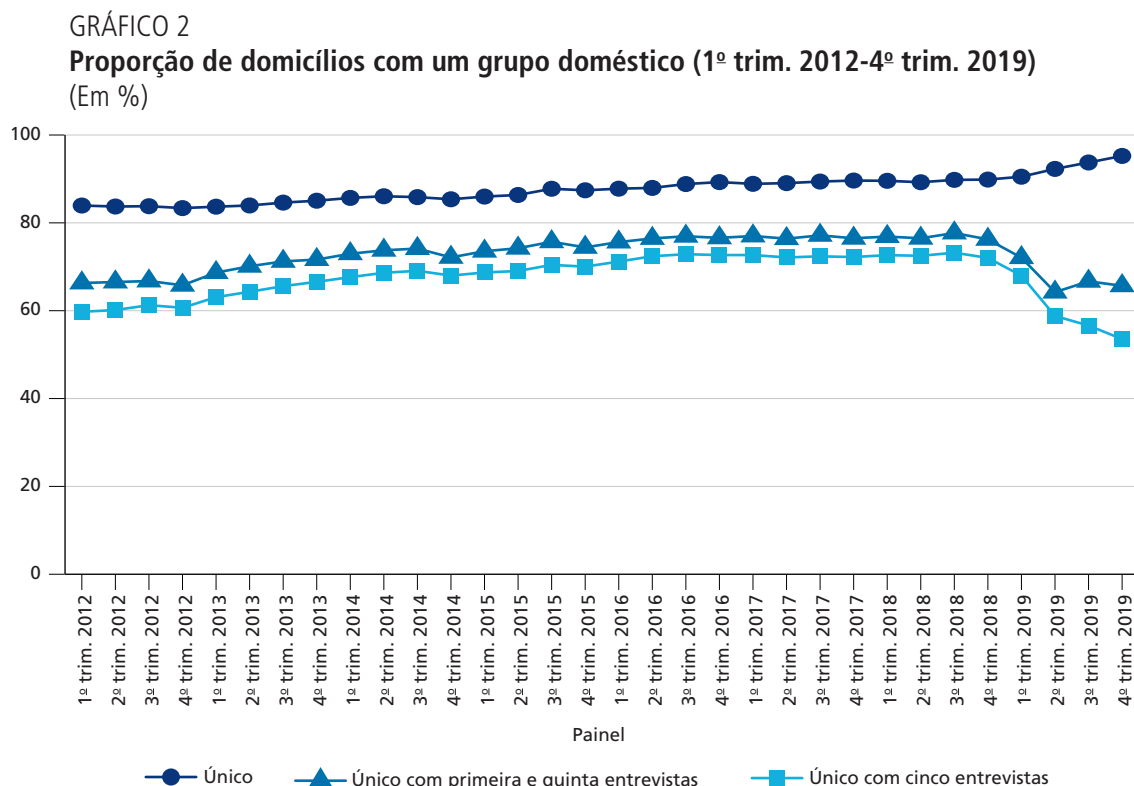
- 3) Ordenar as combinações pela quantidade total de entrevistas e pela quantidade de diferenças, de forma a processar primeiro as combinações com maior número de pessoas (entrevistas) e com menor quantidade de diferenças entre os atributos das pessoas.
- 4) Em ordem, para cada combinação:
 - a) se o conjunto de atributos diferentes de uma combinação pertence ao conjunto de erros permitidos, e se a diferença absoluta entre anos de nascimento é de até três anos, os indivíduos da combinação são agregados e ficam indisponíveis para outras combinações:
 - se a condição no domicílio é constante, o indivíduo resultante recebe o código de classe 6; e
 - se a condição no domicílio varia, o indivíduo resultante recebe o código de classe 7.

Indivíduos agregados no passo 4.a possuem um ou dois atributos com diferenças. Os conjuntos de diferenças permitidas são as combinações de um e dois elementos formadas com os atributos sexo, dia, mês e ano de nascimento: {sexo}; {dia}; {mês}; {ano}; {sexo, dia}; {sexo, mês}; {sexo, ano}; {dia, mês}; {dia, ano}; {mês, ano}. Um indivíduo pode ter uma pessoa que difere em dia e mês de nascimento, e outra que difere em sexo e dia de nascimento, mas não uma com sexo, dia e mês de nascimento diferentes.

A junção de fragmentos de indivíduos resolve um outro problema, relativo a pessoas com data de nascimento ignorada em todas as entrevistas. Suponha uma moradora que é mãe da pessoa responsável, em um painel começado no primeiro trimestre de 2015. Em duas entrevistas, atribuem-lhe 88 anos, e, nas outras três, 90 anos. A imputação do ano de nascimento a partir da idade estimada gera três datas de nascimento distintas: 99/99/1927; 99/99/1925; e 99/99/1926. A diferença de anos de nascimento a faria ser representada por três “indivíduos” sem intersecções dos conjuntos de entrevistas. Comparados, a única diferença entre estes indivíduos é o ano de nascimento, levando à reunião dos fragmentos em um indivíduo classe 6. Se houvesse troca de pessoa responsável e reclassificação como sogra, a junção geraria um indivíduo classe 7. Contudo, se a diferença de anos de nascimento fosse maior que três anos, os fragmentos não seriam unidos, gerando três indivíduos classe 5.

2.6 Resultado da identificação

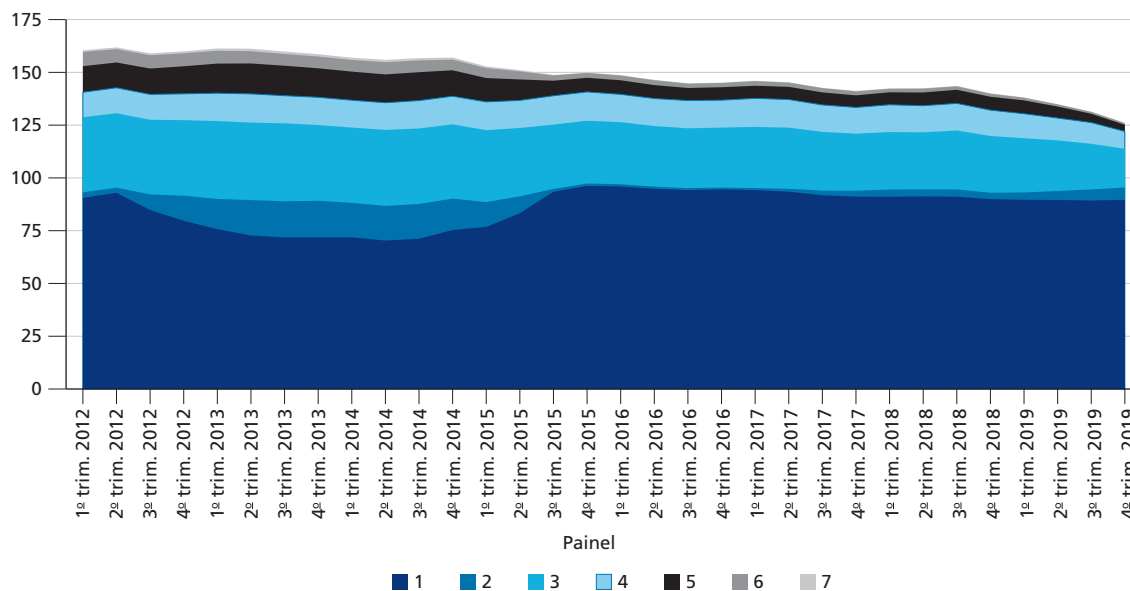
O primeiro resultado relevante da identificação é a quantidade de grupos domésticos. Um domicílio ocupado por mais de um grupo doméstico pode ser usado na análise de dados transversais, mas não em painel. O gráfico 2 exibe, para os painéis de 2012 a 2019, a proporção de domicílios com um grupo doméstico, e com entrevistas nas cinco visitas, ou ao menos nas primeira e quinta visitas. Os painéis de 2019, que possuem entrevistas em 2020, foram bastante afetados pela pandemia, que levou a um aumento considerável das entrevistas não realizadas (IBGE, 2021).



Fonte: Microdados trimestrais da PNAD Contínua (período 2012-2020). Disponível em: <<https://bit.ly/3ILDAfr>>. Elaboração do autor.

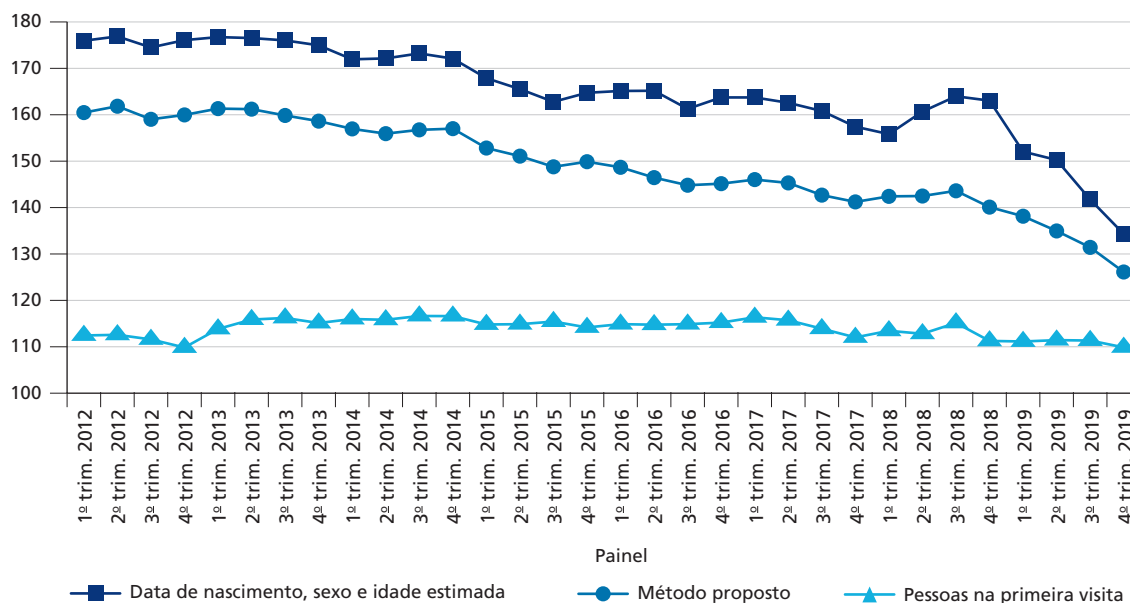
A quantidade de indivíduos identificados segundo a classe de identificação pode ser conferida no gráfico 3. A maior parte dos indivíduos – em torno de 88% nos primeiros painéis, e 94% a partir de 2016 – pertence às classes 1 a 4 de identificação, que são as de menor incerteza. Note-se a predominância de indivíduos classe 1, que são aqueles cujo grupo doméstico (único no domicílio ou não) não muda a composição, e que apresentam o mesmo sexo, data de nascimento, condição no domicílio e número de ordem em todas as entrevistas. A segunda maior classe é a 3, a qual pertencem os indivíduos que aparecem em todas as entrevistas de seu grupo doméstico com mesmo sexo e data de nascimento – apesar das mudanças de composição do grupo. Os indivíduos das classes 5, 6 e 7, as de maior incerteza, e as duas últimas apresentando variações em data de nascimento ou sexo devido a erros de declaração ou registro, apresentam-se sempre em proporção reduzida.

GRÁFICO 3
Quantidade de indivíduos, segundo a classe de identificação (1º trim. 2012-4º trim. 2019)
 (Em 1 mil)



Fonte: Microdados trimestrais da PNAD Contínua (período 2012-2020). Disponível em: <<https://bit.ly/3ILDfFr>>. Elaboração do autor.

GRÁFICO 4
Comparação da quantidade de indivíduos, segundo o método de identificação (1º trim. 2012-4º trim. 2019)
 (Em 1 mil)



Fonte: Microdados trimestrais da PNAD Contínua. Disponível em: <<https://bit.ly/3ILDfFr>>. Elaboração do autor.

Para comparar o método de identificação ora sugerido à identificação feita apenas com sexo, data de nascimento e idade estimada, a referência é a quantidade de indivíduos que existiria se todos os que participaram da primeira entrevista participassem das demais. Se não houvesse atrito e se os grupos domésticos não mudassem de composição, a quantidade de indivíduos seria igual a de registros de pessoa na primeira visita. O gráfico 4 permite comparar a referência e a quantidade de indivíduos obtida pelos dois métodos. A inclinação das curvas que representam os métodos se deve à redução da quantidade de grupos domésticos nos domicílios dos painéis mais recentes (gráfico 2).

3 CONSIDERAÇÕES FINAIS

Este texto apresentou, de forma resumida, um método para identificar os grupos domésticos e os indivíduos nos painéis de domicílios da PNAD Contínua, o que possibilita seu uso. Este método é empregado pelo aplicativo Pynad, mencionado na seção 1, que monta e identifica os painéis e acrescenta às bases de dados um conjunto de variáveis para facilitar o uso.

Contudo, a análise de dados longitudinais exige mais que a montagem e a identificação dos painéis: requer uma série de considerações sobre problemas como o atrito, a consistência dos dados entre entrevistas, o condicionamento dos entrevistados, bem como a expansão da amostra. Existe ampla literatura sobre estas questões, mas não é possível abordar todas no espaço restrito deste estudo.

Brevemente, deve ser publicado um texto para discussão do Ipea descrevendo, com mais detalhes, o método de identificação proposto e discutindo alguns dos problemas relacionados ao uso dos dados em painel da PNAD Contínua, sem, obviamente, a pretensão de dar respostas definitivas. De fato, a forma de lidar com tais problemas depende, em última instância, do objeto de estudo e das técnicas escolhidas para trabalhar os dados.

REFERÊNCIAS

- IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional por Amostra de Domicílios Contínua**: notas metodológicas. Rio de Janeiro: IBGE, 2014. v. 1.
- _____. **Manual básico da entrevista**: Pesquisa Nacional por Amostra de Domicílios Contínua. Rio de Janeiro: IBGE, 2016.
- _____. **Pesquisa Nacional por Amostra de Domicílios Contínua**: Nota Técnica nº 4. Rio de Janeiro: IBGE, 2021.
- RIBAS, R.; SOARES, S. S. D. **Sobre o painel da Pesquisa Mensal de Emprego**. Ipea: Brasília, 2008. (Texto para Discussão, 1348).

Ipea – Instituto de Pesquisa Econômica Aplicada

EDITORIAL

Chefe do Editorial

Aeromilson Trajano de Mesquita

Assistentes da Chefia

Rafael Augusto Ferreira Cardoso

Samuel Elias de Souza

Supervisão

Camilla de Miranda Mariath Gomes

Everson da Silva Moura

Revisão

Alice Souza Lopes

Amanda Ramos Marques

Ana Clara Escórcio Xavier

Clícia Silveira Rodrigues

Luiz Gustavo Campos de Araújo Souza

Olavo Mesquita de Carvalho

Regina Marta de Aguiar

Reginaldo da Silva Domingos

Brena Rolim Peixoto da Silva (estagiária)

Nayane Santos Rodrigues (estagiária)

Editoração

Anderson Silva Reis

Cristiano Ferreira de Araújo

Danielle de Oliveira Ayres

Danilo Leite de Macedo Tavares

Leonardo Hideki Higa

*The manuscripts in languages other than Portuguese
published herein have not been proofread.*

Missão do Ipea

Aprimorar as políticas públicas essenciais ao desenvolvimento brasileiro por meio da produção e disseminação de conhecimentos e da assessoria ao Estado nas suas decisões estratégicas.



ipea Instituto de Pesquisa
Econômica Aplicada

MINISTÉRIO DA
ECONOMIA

