

TEXTO PARA DISCUSSÃO N° 1348

SOBRE O PAINEL DA PESQUISA MENSAL DE EMPREGO (PME) DO IBGE

**Rafael Perez Ribas
Sergei Suarez Dillon Soares**

Rio de Janeiro, agosto de 2008

TEXTO PARA DISCUSSÃO N° 1348

SOBRE O PAINEL DA PESQUISA MENSAL DE EMPREGO (PME) DO IBGE*

Rafael Perez Ribas**

Sergei Suarez Dillon Soares***

Rio de Janeiro, agosto de 2008

* Os autores agradecem os comentários de Carlos Henrique Corseuil e Maurício Cortez Reis. Os erros remanescentes são de responsabilidade dos autores. Todas as sintaxes neste trabalho encontram-se à disposição para todos que as queiram utilizar. Enviar pedidos para <sergei.soares@ipea.gov.br> ou <rafael.ribas@undp-povertycentre.org>.

** Do Centro Internacional de Pobreza.

*** Técnico de Planejamento e Pesquisa da Diretoria de Estudos Sociais do Ipea.

Governo Federal

**Ministro de Estado Extraordinário de
Assuntos Estratégicos** – Roberto Mangabeira Unger

**Secretaria de Assuntos Estratégicos
da Presidência da República**

ipea Instituto de Pesquisa
Econômica Aplicada

Fundação pública vinculada à Secretaria de Assuntos Estratégicos da Presidência da República, o Ipea fornece suporte técnico e institucional às ações governamentais, possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro, e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente

Marcio Pochmann

Diretor de Administração e Finanças

Fernando Ferreira

Diretor de Estudos Macroeconômicos

João Sicsú

Diretor de Estudos Sociais

Jorge Abrahão de Castro

Diretora de Estudos Regionais e Urbanos

Liana Maria da Frota Carleial

Diretor de Estudos Setoriais

Márcio Wohlers de Almeida

Diretor de Cooperação e Desenvolvimento

Mário Lisboa Theodoro

Chefe de Gabinete

Persio Marco Antonio Davison

Assessor-Chefe de Comunicação

Estanislau Maria de Freitas Júnior

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

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

ISSN 1415-4765

JEL: C23, C81

TEXTO PARA DISCUSSÃO

Publicação cujo objetivo é divulgar resultados de estudos desenvolvidos pelo Ipea, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou da Secretaria de Assuntos Estratégicos da Presidência da República.

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

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO 7

2 PESQUISA MENSAL DE EMPREGO 8

3 O PAINEL DA PME 10

4 IMPLICAÇÃO DE MUDANÇAS NO EMPARELHAMENTO SOBRE ALGUNS RESULTADOS 22

5 CONCLUSÃO 24

REFERÊNCIAS 25

APÊNDICE 26

SINOPSE

O objetivo deste texto é fornecer as informações necessárias para que pesquisadores possam utilizar a Pesquisa Mensal de Emprego (PME) e, em particular, o painel de domicílios. O painel da PME é um instrumento extremamente útil de análise que, no entanto, tem sido subutilizado devido a uma série de dificuldades práticas, principalmente na sua (re)construção. Escrevemos este texto para fornecer a outros pesquisadores alguns instrumentos para lidar com estas dificuldades.

O texto tem três partes. A primeira trata da leitura e identificação do painel da PME, inclusive o esquema de rotação da pesquisa. Até esta tarefa relativamente simples torna-se complexa devido a algumas peculiaridades de como a PME é organizada.

Na segunda parte, fornecemos alguns procedimentos para aumentar a taxa de identificação no painel de pessoas. A hipótese por trás destes procedimentos é que o erro na declaração de informações individuais dificulta a reconstituição mais precisa deste painel.

Finalmente, na última parte, apontamos qual é a implicação de uma reconstituição menos precisa do painel de pessoas sobre os resultados de um exemplo de análise do mercado de trabalho.

ABSTRACT

The purpose of this article is to explain how to use the *Pesquisa Mensal de Emprego* (PME) and in particular to explain in detail how to set up the household panel. The PME household panel is an extremely useful analytical instrument that has been sadly underused due to several practical construction difficulties. The main objective of this text is to dispel these difficulties so as to open the household panel to researchers who previously did not know how to set it up.

The first part of the text explains how to read the microdata and set up the PME household panel, and includes an explanation of its rotation scheme. Even this simple task becomes relatively complex due to how the PME is organized.

The second part of the text deals with individuals and supplies some simple procedures for increasing the percentage of individuals successfully identified in the household panel. These procedures work because identifiable reporting error and identifiable changes in household composition reduce the percentage of individuals identified.

Finally, we show how increasing the rate of individual identification improves econometric estimates using as an example labor market insertion.

1 INTRODUÇÃO

A Pesquisa Mensal de Emprego (PME) do Instituto Brasileiro de Geografia e Estatística (IBGE) é uma pesquisa domiciliar de periodicidade mensal, que apresenta um esquema de amostragem igual ao do *US Current Population Survey* (CPS). Suas informações são obtidas de uma amostra probabilística de aproximadamente 40 mil domicílios situados nas Regiões Metropolitanas (RMs) do Rio de Janeiro, de São Paulo, de Porto Alegre, de Belo Horizonte, de Recife e de Salvador. O tema básico da pesquisa é o trabalho, constando algumas características demográficas e educacionais com o objetivo de possibilitar melhor entendimento sobre este tema.

Os microdados da PME podem ser trabalhados tanto na forma empilhada como na forma de série de tempo mensal. Contudo, esta base de dados é muito mais rica, visto que possibilita a estimação de transições, pois contém um painel que acompanha as unidades amostrais por até oito entrevistas. Particularmente, a unidade amostral observada em repetidos meses é o domicílio. Se os indivíduos que residem neste domicílio não mudarem, é possível também acompanhá-los longitudinalmente. Contudo, não é possível observar todos os indivíduos da amostra em todas as oito entrevistas, o que faz da PME um painel incompleto, ou desequilibrado (*unbalanced panel*), do ponto de vista individual.

Basicamente, são duas as causas para o chamado desgaste (ou atrito) do painel: a mobilidade geográfica das pessoas na amostra e a recusa de entrevista. Neste artigo, contudo, levantamos ainda a hipótese de uma terceira fonte de desgaste do painel. Essa fonte estaria relacionada à imprecisão na declaração das informações utilizadas como critério de emparelhamento na reconstituição do painel da PME, por meio dos microdados. A PME não reporta um código que possibilite identificar com certeza a mesma pessoa em períodos distintos. Para contornar esse problema, é comum utilizar algumas características individuais reportadas na pesquisa para identificar a mesma pessoa em duas ou mais entrevistas. Contudo, se alguma destas características for inconsistente entre os períodos, a pessoa nunca será encontrada. Dessa forma, um “falso atrito” de observações pode ser gerado, sobreestimando o verdadeiro desgaste do painel.

O objetivo deste artigo é levantar alguns pontos sobre como construir (ou reconstituir) o painel da PME e testar qual a implicação do “falso atrito” sobre alguns resultados. Para isso, descrevemos como é o esquema de rotação da amostra na pesquisa, como identificar o mesmo domicílio em diferentes entrevistas e, por fim, como lidar com o problema do emparelhamento de pessoas.

Nossos resultados apontam que, com o uso de um algoritmo de emparelhamento mais avançado do que o convencionalmente usado, foi possível reduzir em quase 50% a taxa de atrito nos primeiros intervalos de meses do painel. Nos intervalos maiores, onde o desgaste é maior, a recuperação de casos representa algo próximo de 6 pontos percentuais (p.p.) na antiga pesquisa e de 20 p.p. na nova pesquisa. Além disso, a utilização deste algoritmo pode vir a aumentar tanto a eficiência quanto a consistência de estimativas, principalmente quando o pesquisador trabalhar com uma amostra muito restrita.

2 PESQUISA MENSAL DE EMPREGO

A PME é a pesquisa domiciliar com maior série ininterrupta no Brasil. A primeira PME foi a campo em janeiro de 1980, nas RMs do Rio de Janeiro e de São Paulo, com 12 quesitos de identificação, 4 perguntas sociodemográficas e 17 variáveis sobre emprego, desemprego, e rendimento do trabalho. As RMs de Porto Alegre e de Belo Horizonte foram incorporadas à pesquisa em abril de 1980, enquanto as RMs de Recife e de Salvador ingressaram em junho do mesmo ano.

Do início da pesquisa a janeiro de 1982, a PME era parte integrante da Pesquisa Nacional por Amostra de Domicílios (Pnad), utilizando-se do mesmo desenho amostral. Em fevereiro de 1982, deu início à implantação gradativa de reformulações na amostra, baseadas nas informações do Censo Demográfico de 1980.¹ As reformulações foram concluídas em maio daquele mesmo ano, quando se introduziu também um novo e mais amplo questionário. Esse questionário possuía 16 quesitos de identificação, 6 questões sociodemográficas, 5 sobre educação, e 28 relacionadas ao emprego e desemprego. O questionário permaneceu inalterado até dezembro de 2002, o que gera uma série de 248 meses com a mesma metodologia de pesquisa.

Ao se pretender resultados para cada RM abrangida, separadamente, a PME incorporou aspectos de um plano amostral autoponderado. Dentro de cada RM, a probabilidade de um domicílio ou de uma pessoa qualquer pertencer à amostra é constante e igual à fração amostral. A tabela 1 reporta as frações amostrais das RMs e a respectiva quantidade de domicílios selecionados para a amostra, para a série anterior a 2002. Note-se que, para cerca de 1/4 dos domicílios selecionados para a amostra, a entrevista não era efetivamente realizada.

TABELA 1
Distribuição regional da amostra e das entrevistas nos domicílios da antiga PME

| RMs | Número de setores | Fração | Número de domicílios | Entrevistas realizadas ^a | % |
|----------------|-------------------|--------|----------------------|-------------------------------------|-------|
| Recife | 196 | 1/170 | 5.022 | 3.965 | 78,95 |
| Salvador | 169 | 1/170 | 5.100 | 3.912 | 76,71 |
| Belo Horizonte | 244 | 1/170 | 7.019 | 5.643 | 80,40 |
| Rio de Janeiro | 315 | 1/430 | 7.826 | 5.792 | 74,01 |
| São Paulo | 332 | 1/600 | 8.366 | 6.001 | 71,73 |
| Porto Alegre | 254 | 1/170 | 6.762 | 5.310 | 78,53 |
| Total | 1.510 | - | 40.095 | 30.623 | 76,38 |

Fonte: IBGE (1998).

^a Valores referentes ao mês de junho de 1998.

1. Em agosto de 1988, efetuou-se nova alteração no desenho da amostra, que resultou numa redução do número de unidades selecionadas, diminuindo os custos operacionais, mas mantendo a precisão dos resultados finais em níveis aceitáveis. Em outubro de 1993, a PME começou a implantar gradativamente uma nova amostra com base no Censo Demográfico de 1991, concluindo essa implantação em janeiro de 1994.

De acordo com IBGE (2002), as recentes mudanças na estrutura produtiva, na alocação da mão-de-obra e nas relações de trabalho, juntamente com as novas recomendações da Organização Internacional do Trabalho (OIT) para a investigação da força de trabalho, levaram o instituto a realizar uma revisão da PME em todos os seus aspectos, metodológicos e processuais. Assim, em março de 2002, uma nova PME, com um questionário ainda mais amplo, foi a campo com 14 quesitos de identificação e 6 questões sociodemográficas,² 14 sobre educação, 67 sobre emprego e desemprego. Os principais objetivos da revisão foram: implementação de algumas mudanças conceituais no tema trabalho; ampliação da investigação para se ter melhor conhecimento da População Economicamente Ativa (PEA) e da População em Idade Ativa (PIA); e melhor operacionalização dos quesitos para captação das informações de forma a aprimorar a mensuração dos fenômenos.

Note-se que ambas as pesquisas, com a antiga e a nova metodologia, foram a campo concomitantemente por dez meses, o que, em princípio, permite comparar as mudanças ocorridas. A nova PME deverá permanecer inalterada, pelo menos, até dezembro de 2010, quando o IBGE a substituirá pela mais completa Pesquisa Domiciliar Contínua. Isto levará a uma série de 106 observações no tempo.

Em relação ao desenho amostral, na tabela 2, aumentou-se em quase 500 o número de setores selecionados e calculou-se uma nova fração amostral, que diminuiu a probabilidade de um domicílio qualquer pertencer à amostra. Por conseqüência, a nova pesquisa reduziu em quase 3 mil casos o número de unidades domiciliares selecionadas. Contudo, o percentual de entrevistas realizadas manteve-se igual. O IBGE (2002) recomenda que alguns cuidados devem ser tomados ao compararmos os resultados da nova pesquisa com os da antiga. Um deles diz respeito justamente à composição da amostra por RM.

TABELA 2
Distribuição regional da amostra e das entrevistas nos domicílios da nova PME

| RMs | Número de setores | Fração | Número de domicílios | Entrevistas realizadas ^a | % |
|----------------|-------------------|--------|----------------------|-------------------------------------|-------|
| Recife | 261 | 1/200 | 4.715 | 3.714 | 78,77 |
| Salvador | 243 | 1/200 | 4.684 | 3.260 | 69,60 |
| Belo Horizonte | 359 | 1/200 | 6.644 | 5.253 | 79,06 |
| Rio de Janeiro | 406 | 1/500 | 7.576 | 5.339 | 70,47 |
| São Paulo | 431 | 1/700 | 7.820 | 6.276 | 80,26 |
| Porto Alegre | 329 | 1/200 | 5.773 | 4.470 | 77,43 |
| Total | 2.029 | - | 37.212 | 28.312 | 76,08 |

Fonte: IBGE (2002).

^a Valores referentes ao mês de junho de 2002.

Por fim, uma outra alteração ocorrida na nova pesquisa foi na forma de leitura dos microdados. Os microdados da PME são disponibilizados pelo IBGE em

2. Ao contrário da antiga pesquisa, que só entrevistava pessoas com dez anos ou mais de idade, as novas questões sociodemográficas são perguntadas a todos os indivíduos da unidade domiciliar. Na prática, isso possibilita uma reconstituição mais precisa da composição demográfica domiciliar.

arquivos de texto no formato *American Standard Code for Information Interchange* (ASCII). Para a antiga PME, os arquivos ASCII se apresentam com “pessoas na linha do domicílio”. Esta forma de apresentação consiste em ordenar as informações de pessoas depois das informações do domicílio em que habitam. Conforme o diagrama 1, por exemplo, o primeiro domicílio, identificado pela variável I_1 e descrito pelas variáveis D_1^k , possui duas pessoas residentes, com características $P_{1,1}^k$ e $P_{1,2}^k$. O segundo domicílio, identificado pela variável I_2 e descrito pelas variáveis D_2^k , possui três residentes, descritos por $P_{2,1}^k$, $P_{2,2}^k$ e $P_{2,3}^k$.

DIAGRAMA 1
Informações de domicílio na linha das pessoas

| | | | | | | | | | | | | |
|-------|---------|---------|---------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| I_1 | D_1^1 | D_1^2 | D_1^3 | $P_{1,1}^1$ | $P_{1,1}^2$ | $P_{1,1}^3$ | $P_{1,2}^1$ | $P_{1,2}^2$ | $P_{1,2}^3$ | | | |
| I_2 | D_2^1 | D_2^2 | D_2^3 | $P_{2,1}^1$ | $P_{2,1}^2$ | $P_{2,1}^3$ | $P_{2,2}^1$ | $P_{2,2}^2$ | $P_{2,2}^3$ | $P_{2,3}^1$ | $P_{2,3}^2$ | $P_{2,3}^3$ |

Na nova PME, os arquivos ASCII se apresentam com o “domicílio na linha das pessoas”. Este formato consiste em repetir, para cada pessoa, todas as informações do seu domicílio, conforme mostra o diagrama 2.

DIAGRAMA 2
Informações de domicílio na linha das pessoas

| | | | | | | |
|-------|---------|---------|---------|-------------|-------------|-------------|
| I_1 | D_1^1 | D_1^2 | D_1^3 | $P_{1,1}^1$ | $P_{1,1}^2$ | $P_{1,1}^3$ |
| I_1 | D_1^1 | D_1^2 | D_1^3 | $P_{1,2}^1$ | $P_{1,2}^2$ | $P_{1,2}^3$ |
| I_2 | D_2^1 | D_2^2 | D_2^3 | $P_{2,1}^1$ | $P_{2,1}^2$ | $P_{2,1}^3$ |
| I_2 | D_2^1 | D_2^2 | D_2^3 | $P_{2,2}^1$ | $P_{2,2}^2$ | $P_{2,2}^3$ |
| I_2 | D_2^1 | D_2^2 | D_2^3 | $P_{2,3}^1$ | $P_{2,3}^2$ | $P_{2,3}^3$ |

A leitura dos microdados, particularmente na antiga PME, pode ser penosa para os não-iniciados. Felizmente, a partir de 2001, o IBGE passou a disponibilizar os microdados da antiga PME desde 1991, em um formato tal como o do diagrama 2.

3 O PAINEL DA PME

Na PME, de modo a se ter maior segurança nas comparações mensais dos resultados sobre o mercado de trabalho, os mesmos informantes deveriam ser entrevistados ao longo do tempo. Somente dessa forma haveria a garantia de que as variações verificadas não estariam sendo provocadas pela troca de informantes. Contudo, para minimizar o cansaço imposto aos informantes de serem entrevistados por repetidas vezes, a pesquisa adota um esquema de rotação de painéis (*rotating panel*). Isto quer dizer que os domicílios não são entrevistados durante todos os meses de pesquisa, mas eles entram e saem da amostra de acordo com um padrão predefinido.

A reconstituição deste painel torna a pesquisa ainda mais rica em termos de possibilidades de estudos. No entanto, alguns cuidados devem ser tomados no

processo de reconstituição. Detalhes sobre a estrutura e a identificação de pessoas e domicílios no painel são colocados a seguir.

Na antiga PME, um painel equivale a um conjunto de domicílios selecionados que é dividido em quatro grupos rotacionais correspondentes, cada qual, a uma remessa de setores entrevistados em uma semana específica do mês. Os grupos rotacionais são indicados por uma letra, que identifica o painel, acompanhada de um subscrito correspondente à semana do mês. Se no mês t for aplicado o painel B (B1, B2, B3, B4), no mês $t + 1$ serão aplicados apenas 75% do seu todo (B1, B2, B3), entrando 1/4 do painel seguinte C (C4), e assim sucessivamente. Assim, há a garantia de que 75% dos domicílios são comuns em dois meses consecutivos.

O esquema de rotação, chamado 4-8-4, determinava que, de outubro de um ano ímpar a setembro do ano seguinte, todo mês um grupo de domicílios entraria na pesquisa e seria entrevistado por quatro meses consecutivos. Do quinto ao décimo segundo mês, este grupo sairia da amostra, retornando no décimo terceiro mês e sendo entrevistado por mais quatro vezes. Os domicílios saíam da amostra definitivamente 16 meses depois de sua entrada. É importante salientar que, de outubro de um ano par a setembro do ano seguinte, nenhum grupo novo de domicílios entrava na amostra. Neste período, a cada mês, apenas retornavam os grupos entrevistados que estavam no intervalo de oito meses sem entrevista. Por consequência, a cada par de anos, 100% da amostra se repetem. O quadro 1, feito para os anos 1996 a 1999, facilita a compreensão do esquema de rotação.

QUADRO 1
Painel rotativo da antiga PME

| Mês/ano | 1996 | | | | 1997 | | | | 1998 | | | | 1999 | | | |
|-----------|-----------|----|----|----|-----------|----|----|----|------|----|----|----|------|----|----|----|
| Janeiro | D1 | D2 | D3 | D4 | D1 | D2 | D3 | D4 | G1 | G2 | G3 | G4 | G1 | G2 | G3 | G4 |
| Fevereiro | D1 | D2 | D3 | E4 | D1 | D2 | D3 | E4 | G1 | G2 | G3 | H4 | G1 | G2 | G3 | H4 |
| Março | D1 | D2 | E3 | E4 | D1 | D2 | E3 | E4 | G1 | G2 | H3 | H4 | G1 | G2 | H3 | H4 |
| Abril | D1 | E2 | E3 | E4 | D1 | E2 | E3 | E4 | G1 | H2 | H3 | H4 | G1 | H2 | H3 | H4 |
| Maiο | E1 | E2 | E3 | E4 | E1 | E2 | E3 | E4 | H1 | H2 | H3 | H4 | H1 | H2 | H3 | H4 |
| Junho | E1 | E2 | E3 | F4 | E1 | E2 | E3 | F4 | H1 | H2 | H3 | I4 | H1 | H2 | H3 | I4 |
| Julho | E1 | E2 | F3 | F4 | E1 | E2 | F3 | F4 | H1 | H2 | I3 | I4 | H1 | H2 | I3 | I4 |
| Agosto | E1 | F2 | F3 | F4 | E1 | F2 | F3 | F4 | H1 | I2 | I3 | I4 | H1 | I2 | I3 | I4 |
| Setembro | F1 | F2 | F3 | F4 | F1 | F2 | F3 | F4 | I1 | I2 | I3 | I4 | I1 | I2 | I3 | I4 |
| Outubro | F1 | F2 | F3 | D4 | F1 | F2 | F3 | G4 | I1 | I2 | I3 | G4 | I1 | I2 | I3 | J4 |
| Novembro | F1 | F2 | D3 | D4 | F1 | F2 | G3 | G4 | I1 | I2 | G3 | G4 | I1 | I2 | J3 | J4 |
| Dezembro | F1 | D2 | D3 | D4 | F1 | G2 | G3 | G4 | I1 | G2 | G3 | G4 | I1 | J2 | J3 | J4 |

Fonte: Lopes (2002).

Por exemplo, o grupo rotacional D1 entrou em janeiro de 1996, permaneceu até abril, ficou de maio até dezembro fora, voltou em janeiro de 1997 e, após abril de 1997, saiu definitivamente da pesquisa. Cabe notar que, de anos pares para anos ímpares, de janeiro a setembro, exatamente os mesmos domicílios compõem a

amostra. Isto pode ser visto comparando, por exemplo, junho de 1996 com o mesmo mês em 1997. Nos meses de outubro a dezembro, alguns grupos rotacionais mudam. De um ano ímpar para um ano par, os grupos de domicílios são todos diferentes, o que pode ser verificado de junho de 1997 a junho de 1998. Mais uma vez, a exceção são os grupos de domicílios que entram na amostra entre os meses de outubro e dezembro. Conhecer este esquema de rotação evita que o pesquisador gaste muito tempo procurando grupos rotacionais que já saíram da amostra.

No caso da nova PME, o padrão 4-8-4 foi mantido, mas houve um ajustamento no processo de rotação para dar mais condições de acompanhamento longitudinal dos resultados. Os grupos rotacionais foram sincronizados de forma que não acontecesse o chamado *blackout* a cada dois anos, quando toda a amostra era trocada. Desse modo, aumentou-se o número de grupos rotacionais de quatro para oito, rodando dois grupos por mês em vez de apenas um, como acontecia. O quadro 2 ilustra o novo sistema.

QUADRO 2
Painel rotativo da nova PME

| Mês/ano | 2003 | | | | | | | | 2004 | | | | | | | | 2005 | | | | | | | |
|-----------|-----------|-----------|-----------|-----------|-----------|----|----|----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----|----|----|----|-----------|-----------|-----------|
| Janeiro | C1 | C2 | C3 | C4 | D5 | D6 | D7 | D8 | F1 | F2 | F3 | F4 | D5 | D6 | D7 | D8 | F1 | F2 | F3 | F4 | G5 | G6 | G7 | G8 |
| Fevereiro | E1 | C2 | C3 | C4 | C5 | D6 | D7 | D8 | E1 | F2 | F3 | F4 | F5 | D6 | D7 | D8 | H1 | F2 | F3 | F4 | F5 | G6 | G7 | G8 |
| Março | E1 | E2 | C3 | C4 | C5 | C6 | D7 | D8 | E1 | E2 | F3 | F4 | F5 | F6 | D7 | D8 | H1 | H2 | F3 | F4 | F5 | F6 | G7 | G8 |
| Abril | E1 | E2 | E3 | C4 | C5 | C6 | C7 | D8 | E1 | E2 | E3 | F4 | F5 | F6 | F7 | D8 | H1 | H2 | H3 | F4 | F5 | F6 | F7 | G8 |
| Mai | E1 | E2 | E3 | E4 | C5 | C6 | C7 | C8 | E1 | E2 | E3 | E4 | F5 | F6 | F7 | F8 | H1 | H2 | H3 | H4 | F5 | F6 | F7 | F8 |
| Junho | D1 | E2 | E3 | E4 | E5 | C6 | C7 | C8 | G1 | E2 | E3 | E4 | E5 | F6 | F7 | F8 | G1 | H2 | H3 | H4 | H5 | F6 | F7 | F8 |
| Julho | D1 | D2 | E3 | E4 | E5 | E6 | C7 | C8 | G1 | G2 | E3 | E4 | E5 | E6 | F7 | F8 | G1 | G2 | H3 | H4 | H5 | H6 | F7 | F8 |
| Agosto | D1 | D2 | D3 | E4 | E5 | E6 | E7 | C8 | G1 | G2 | G3 | E4 | E5 | E6 | E7 | F8 | G1 | G2 | G3 | H4 | H5 | H6 | H7 | F8 |
| Setembro | D1 | D2 | D3 | D4 | E5 | E6 | E7 | E8 | G1 | G2 | G3 | G4 | E5 | E6 | E7 | E8 | G1 | G2 | G3 | G4 | H5 | H6 | H7 | H8 |
| Outubro | F1 | D2 | D3 | D4 | D5 | E6 | E7 | E8 | F1 | G2 | G3 | G4 | G5 | E6 | E7 | E8 | I1 | G2 | G3 | G4 | G5 | H6 | H7 | H8 |
| Novembro | F1 | F2 | D3 | D4 | D5 | D6 | E7 | E8 | F1 | F2 | G3 | G4 | G5 | G6 | E7 | E8 | I1 | I2 | G3 | G4 | G5 | G6 | H7 | H8 |
| Dezembro | F1 | F2 | F3 | D4 | D5 | D6 | D7 | E8 | F1 | F2 | F3 | G4 | G5 | G6 | G7 | E8 | I1 | I2 | I3 | G4 | G5 | G6 | G7 | H8 |

Fonte: IBGE (2002).

Na nova metodologia, o grupo E1 em negrito, por exemplo, é entrevistado de fevereiro a maio de 2003 (quatro meses) e novamente de fevereiro a maio de 2004. A principal mudança é que a cada mês dois grupos, cada um equivalente a 1/8 da amostra, saem. Isto fica claro de janeiro a fevereiro de 2003, período quando os grupos C1 e D5 são trocados pelo E1 e C5. Ou seja, a sobreposição de 75% da amostra de um mês para o outro foi mantida, mas com a rotação de dois grupos de 1/8 cada e não de um único grupo de 1/4. Como resultado, a cada 12 meses, metade da amostra é sempre comum. Repare que os meses de junho em 2003 e em 2004 compartilham os grupos E2, E3, E4 e E5. Da mesma forma, os meses de junho em

2004 e em 2005 compartilham os grupos G1, F6, F7 e F8. Não há dúvida de que o esquema atual de rotação amplia a possibilidade de se investigar certos fenômenos sociais e econômicos que ocorrem entre um ano e outro.

Conhecidos os esquemas de rotação das PMEs, o primeiro passo para reconstituir o painel é saber como identificar o mesmo domicílio em dois períodos. As variáveis que identificam o domicílio na pesquisa são as seguintes:

| Antiga | Descrição | Nova | Descrição |
|--------|----------------------|------|----------------------|
| v010 | Unidade da federação | v035 | Região Metropolitana |
| v101 | Número no 202/203 | v040 | Número de controle |
| v102 | Número de controle | v050 | Número de série |
| v103 | Número de série | v060 | Painel |
| v106 | Remessa | v063 | Grupo rotacional |

Para identificar a entrevista de cada domicílio univocamente, basta acrescentar ainda a variável de “mês da pesquisa” (v070 na nova e v105 na antiga) ou a variável de “número da pesquisa no domicílio” (v072 na nova). A variável de “número da pesquisa no domicílio”, assim como a variável “painel” (que na nova PME é representada por v060), não consta nas bases de dados da antiga PME. Estas variáveis são muito úteis na reconstituição do painel e podem ser construídas a partir das demais variáveis de identificação de acordo com a necessidade.

A tabela 3 apresenta a sobreposição da amostra nas duas PMEs. Para os domicílios, essa sobreposição é calculada a partir das variáveis descritas acima. Para os indivíduos, o emparelhamento é mais complicado. A PME não reporta um código que possibilite identificar com certeza a mesma pessoa em períodos distintos.³ Além disso, a pesquisa não coleta informações sobre números de registros (NIS, RG, CPF, título de eleitor etc.) e nem divulga os nomes das pessoas entrevistadas, por motivos óbvios de confidencialidade.

De acordo com Lopes (2002), as informações disponíveis para o emparelhamento de pessoas são dia, mês e ano de nascimento, e sexo. Denominamos esta combinação variáveis de “emparelhamento básico”. Somada a essas características, alguns autores utilizam ainda a escolaridade, a condição no domicílio e/ou o número de ordem da pessoa como critérios de emparelhamento. Porém, essas características são sensíveis a mudanças no tempo, já que as pessoas nascem, morrem ou se mudam e, por conseqüência, alteram a composição domiciliar. Além disso, elas não acrescentam muito rigor ao emparelhamento, pois, com exceção dos gêmeos, é muito difícil encontrar, em um mesmo domicílio, pessoas com a mesma data de nascimento.

3. Apesar de alguns pensarem que o número de ordem da pessoa no domicílio não muda entre as entrevistas, esta variável não possui o propósito de identificador no painel. Este número é atribuído independentemente a cada mês de acordo com a condição das pessoas no domicílio naquele instante. Basta uma pessoa no domicílio alterar a sua condição, que o número de ordem dela e das demais não será o mesmo entre duas entrevistas.

Deve-se ressaltar que se, durante o período em que o domicílio permanece na amostra, a família mudar de endereço e outra família passar a ocupar a unidade domiciliar, a informação será obtida com a nova família pelo período restante. Portanto, é possível que a sobreposição de domicílio no painel não implique sobreposição de famílias ou pessoas.

TABELA 3
Sobreposição da amostra na PME entre pares de meses seguindo os critérios básicos de emparelhamento
(Em %)

| Intervalo de meses | Antiga PME | | | Nova PME | | |
|--------------------|-------------------|------------|------------|----------|------------|------------|
| | Máximo | Domicílios | Indivíduos | Máximo | Domicílios | Indivíduos |
| 1 | 75,0 | 71,6 | 65,1 | 75,0 | 72,0 | 65,2 |
| 2 | 50,0 | 47,0 | 42,0 | 50,0 | 47,3 | 41,0 |
| 3 | 25,0 | 23,1 | 20,2 | 25,0 | 23,2 | 19,0 |
| 4-8 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 |
| 9 | 25,0 ^a | 22,8 | 19,8 | 12,5 | 11,1 | 5,6 |
| 10 | 50,0 ^a | 45,5 | 38,9 | 25,0 | 22,4 | 11,4 |
| 11 | 75,0 ^a | 67,9 | 57,1 | 37,5 | 33,6 | 17,0 |
| 12 | 100 ^a | 89,7 | 74,1 | 50,0 | 44,5 | 22,3 |
| 13 | 75,0* | 67,0 | 54,5 | 37,5 | 30,5 | 15,2 |
| 14 | 50,0* | 44,3 | 35,4 | 25,0 | 18,3 | 9,0 |
| 15 | 25,0* | 21,8 | 17,1 | 12,5 | 8,1 | 3,9 |
| 16 ou mais | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 |

Fonte: Elaboração própria a partir da PME de 1998, 1999, 2003 e 2004.

^a De ano par para ímpar. Valores referentes à média das amostras presentes nos meses de janeiro a setembro de 1998, para a antiga PME, e nos meses de janeiro a dezembro de 2003, para a nova PME.

Em decorrência dos domicílios que, em períodos subsequentes, passam a ser inexistentes, estão fechados ou cujos habitantes se recusaram a responder à pesquisa, o emparelhamento dos domicílios não é perfeito. Como a identificação do domicílio é precisa, nada se pode fazer em relação a essa forma de desgaste no painel. Se o domicílio não foi entrevistado, nenhum procedimento estatístico pode revelar sua informação. A segunda forma de desgaste – pessoas não emparelhadas porque seu domicílio não foi entrevistado ou porque elas não faziam parte do domicílio no momento da entrevista – é obviamente maior. Contudo, a informação utilizada para o emparelhamento de indivíduos pode não ser tão precisa quanto a informação utilizada no emparelhamento de domicílios e, portanto, a sobreposição de pessoas pode estar subestimada, desde que haja algum erro na informação individual reportada em algum dos meses.

A tabela 4 mostra a taxa de atrito da amostra, que é a proporção do desgaste ocorrido entre os meses sobre a perda máxima de domicílios imposta pelo esquema de rotação. Em ambas as PMEs, as taxas de atrito no intervalo de um mês são em torno de 4% para domicílios e de 13% para indivíduos, sendo crescentes com o aumento

no intervalo dos meses. Até o intervalo de 12 meses, a perda proporcional de domicílio continua semelhante entre as PMEs, chegando a perto de 10%. Porém, as perdas relativas de indivíduos são muito mais altas na PME nova, a partir dos nove meses de intervalo, com mais de 50% de perda, chegando a quase 70% no 16º mês. Na antiga PME, a taxa de atrito de indivíduos varia de 20% a 30%, a partir dos nove meses de intervalo.

TABELA 4
Perdas entre pares de meses como proporção do limite rotacional da PME, seguindo os critérios básicos de emparelhamento
(Em %)

| Intervalo de meses | Antiga PME | | | Nova PME | | |
|--------------------|------------|------------|----------------------------|------------|------------|----------------------------|
| | Domicílios | Indivíduos | | Domicílios | Indivíduos | |
| | | Total | Em domicílios emparelhados | | Total | Em domicílios emparelhados |
| 1 | 4,5 | 13,2 | 9,4 | 4,0 | 13,0 | 9,6 |
| 2 | 6,0 | 16,1 | 11,0 | 5,5 | 18,0 | 13,1 |
| 3 | 7,7 | 19,3 | 12,7 | 7,3 | 23,1 | 16,4 |
| 9 | 8,6 | 20,8 | 13,0 | 11,1 | 54,9 | 45,1 |
| 10 | 9,0 | 22,3 | 14,4 | 10,5 | 54,5 | 45,4 |
| 11 | 9,5 | 23,9 | 15,6 | 10,4 | 54,7 | 45,6 |
| 12 | 10,3 | 25,9 | 16,9 | 11,0 | 55,5 | 45,8 |
| 13 | 10,7 | 27,3 | 18,1 | 18,7 | 59,6 | 42,3 |
| 14 | 11,5 | 29,3 | 19,3 | 26,9 | 64,1 | 38,6 |
| 15 | 13,0 | 31,6 | 20,4 | 35,4 | 68,9 | 34,8 |

Fonte: Elaboração própria a partir da PME de 1998, 1999, 2003 e 2004.

Nota: Valores referentes à média das amostras presentes nos meses de janeiro a setembro de 1998, para a antiga PME, e nos meses de janeiro a dezembro de 2003, para a nova PME.

Uma diferença clara entre o painel da antiga e da nova PME é a razão entre as taxas de atrito de domicílios e indivíduos. Na antiga, o desgaste de indivíduos é duas vezes e meia maior que o desgaste de domicílios, enquanto na nova, o desgaste de indivíduos é de três a cinco vezes maior que o desgaste de domicílios. O aumento relativo no desgaste de indivíduos pode ser tanto relacionado à mudança na metodologia (tamanho e composição da amostra) e/ou à queda na qualidade da informação utilizada como critério para o emparelhamento quanto a mudanças conjunturais e nos arranjos domiciliares. Pela tabela 4, constatamos que entre 65% e 70% dos indivíduos atritados estão em domicílios emparelhados na antiga pesquisa, enquanto na nova pesquisa essa proporção está acima de 80%, dependendo do intervalo de meses.

3.1 OS PROBLEMAS NO EMPARELHAMENTO DE PESSOAS E COMO SOLUCIONÁ-LOS

Como já foi colocado, se a informação utilizada para o emparelhamento de indivíduos não for precisa, a sobreposição de pessoas na amostra pode ser

subestimada. Apresentamos, a seguir, alguns destes exemplos, encontrados aleatoriamente nas PMEs de 2002 e 2003. No primeiro caso, a única pessoa residente no domicílio identificado por '26000012 1 D 1' (v040 + v050 + v060 + v063) apresentava-se na sua primeira entrevista, em junho de 2002, da seguinte forma:

| v203 | v204 | v214 | v224 | v234 | v208 | vdae1 |
|------|------|------|------|------|------|-------|
| 2 | 99 | 99 | 9999 | 65 | 1 | 4 |

Ela era uma mulher (v203 = 2), branca (v208 = 1), de 65 anos (v234 = 65) e com escolaridade entre 8 e 10 anos de estudo completos (vdae1 = 4). Note-se que, o dia (v204), o mês (v214) e o ano de nascimento (v224) não foram informados. No mês seguinte, possivelmente a mesma mulher se apresenta da seguinte forma:

| v203 | v204 | v214 | v224 | v234 | v208 | vdae1 |
|------|------|------|------|------|------|-------|
| 2 | 20 | 5 | 1934 | 68 | 1 | 5 |

Desta vez, ela declarou a sua data de nascimento. Pelo critério básico de emparelhamento, diríamos que elas não são a mesma pessoa, justamente porque uma das datas de nascimento não foi declarada. Além disso, se fosse utilizada outra variável como critério, como idade presumida (v234)⁴ ou escolaridade (vdae1), o emparelhamento também não iria funcionar. Mais um mês à frente, a encontramos novamente, porém com um novo nível de escolaridade:

| v203 | v204 | v214 | v224 | v234 | v208 | vdae1 |
|------|------|------|------|------|------|-------|
| 2 | 20 | 5 | 1934 | 68 | 1 | 3 |

Ou seja, em um mês, essa mulher passou de um nível de escolaridade com mais de 11 anos de estudo completos para menos de 7 anos de estudo completos.

O segundo caso, referente ao domicílio identificado por '26000012 3 D 1', é um exemplo de mudança na composição da família combinada com algumas inconsistências nas informações declaradas. Em junho de 2002, o domicílio se apresentava da seguinte forma:

| v201 | v203 | v205 | v206 | v207 | v208 | v204 | v214 | v224 | v234 |
|------|------|------|------|------|------|------|------|------|------|
| 1 | 2 | 1 | 1 | 1 | 1 | 24 | 12 | 1970 | 31 |
| 2 | 1 | 3 | 3 | 1 | 4 | 7 | 2 | 1990 | 12 |
| 3 | 1 | 3 | 3 | 1 | 4 | 99 | 9 | 1992 | 9 |
| 4 | 2 | 5 | 1 | 2 | 4 | 12 | 6 | 1980 | 21 |
| 5 | 1 | 4 | 3 | 2 | 4 | 2 | 4 | 1998 | 4 |
| 6 | 1 | 4 | 3 | 2 | 4 | 99 | 99 | 1999 | 2 |

4. A idade presumida é uma variável calculada a partir da diferença entre a data de nascimento da pessoa e a data de entrevista. Porém, caso o informante não saiba o ano ou o mês de nascimento da pessoa, o entrevistador pergunta qual é a idade que o informante presume que esta pessoa tenha. Por se tratar de uma variável que exige um menor rigor em sua precisão, a probabilidade de erro na idade presumida é menor do que na data de nascimento.

Eram duas mulheres, provavelmente irmãs, com dois filhos homens cada uma. Em julho daquele ano, uma nova pessoa apareceu no domicílio, provavelmente o irmão das mulheres de acordo com a relação de parentesco (v205):

| v201 | v203 | v205 | v206 | v207 | v208 | v204 | v214 | v224 | v234 |
|------|------|------|------|------|------|------|------|------|------|
| 1 | 2 | 1 | 1 | 1 | 4 | 24 | 12 | 1970 | 31 |
| 2 | 1 | 3 | 3 | 1 | 4 | 7 | 1 | 1990 | 12 |
| 3 | 2 | 3 | 3 | 1 | 4 | 14 | 7 | 1992 | 9 |
| 4 | 2 | 4 | 1 | 2 | 4 | 12 | 6 | 1980 | 22 |
| 5 | 1 | 4 | 3 | 2 | 4 | 2 | 4 | 1998 | 4 |
| 6 | 1 | 4 | 3 | 2 | 4 | 5 | 5 | 1999 | 3 |
| 7 | 1 | 4 | 4 | 1 | 4 | 13 | 12 | 1968 | 33 |

Note-se que a chefe (v201 = 1) passou de branca (v208 = 1) a parda (v208 = 4), um de seus filhos (v201 = 2) mudou o mês de nascimento (v214) e o outro (v201 = 3) mudou de sexo (v203), além de passar a responder o dia e o mês de nascimento. O critério básico de emparelhamento, neste caso, levaria à conclusão de que a chefe perdeu um de seus filhos homens de um mês para o outro, mas ganhou uma filha mulher. A entrada do novo membro no domicílio fez com que a ordem das pessoas entrevistadas (v201) mudasse no mês seguinte:

| v201 | v203 | v205 | v206 | v207 | v208 | v204 | v214 | v224 | v234 |
|------|------|------|------|------|------|------|------|------|------|
| 1 | 2 | 1 | 1 | 1 | 4 | 24 | 12 | 1970 | 31 |
| 2 | 1 | 3 | 3 | 1 | 4 | 7 | 1 | 1990 | 12 |
| 3 | 2 | 4 | 1 | 2 | 4 | 12 | 6 | 1980 | 22 |
| 4 | 1 | 4 | 4 | 1 | 4 | 13 | 12 | 1968 | 33 |
| 5 | 1 | 4 | 3 | 2 | 4 | 2 | 4 | 1998 | 4 |
| 6 | 1 | 4 | 3 | 2 | 4 | 5 | 5 | 1999 | 3 |
| 7 | 1 | 3 | 3 | 1 | 4 | 16 | 8 | 1992 | 9 |

A pessoa v201 = 3 passou para v201 = 7, a v201 = 4 passou para v201 = 3 e a v201 = 7 passou para v201 = 4. Além disso, a criança que mudou de sexo no mês anterior, trocou o sexo e a data de nascimento novamente. Esta família foi observada neste domicílio até a quarta entrevista. Na quinta entrevista, depois dos oito meses de intervalo, ninguém foi encontrado no domicílio. Na sexta entrevista, havia apenas um homem de 22 anos residindo no local e, na oitava e última entrevista, já havia uma outra família de três pessoas.

O último exemplo é de uma família que foi entrevistada as oito vezes, residente no domicílio '26000012 4 D 1'. Essa família é formada por um casal com dois filhos, sendo que um deles possui uma esposa e um filho. Note-se como as datas de

nascimento (v204, v214 e v224) e os níveis de escolaridade (vdae1) do chefe e do cônjuge neste domicílio mudam de um mês para o outro:

| v070 | v072 | v075 | v203 | v205 | v204 | v214 | v224 | v234 | vdae1 |
|------|------|------|------|------|------|------|------|------|-------|
| 6 | 1 | 2002 | 2 | 1 | 5 | 11 | 1959 | 42 | 3 |
| 6 | 1 | 2002 | 1 | 2 | 99 | 99 | 9999 | 65 | 5 |
| 7 | 2 | 2002 | 2 | 1 | 5 | 12 | 1956 | 45 | 3 |
| 7 | 2 | 2002 | 1 | 2 | 1 | 1 | 1939 | 63 | 4 |
| 8 | 3 | 2002 | 2 | 1 | 5 | 12 | 1956 | 45 | 3 |
| 8 | 3 | 2002 | 1 | 2 | 1 | 1 | 1939 | 63 | 5 |
| 9 | 4 | 2002 | 2 | 1 | 5 | 12 | 1956 | 45 | 3 |
| 9 | 4 | 2002 | 1 | 2 | 1 | 1 | 1939 | 63 | 2 |
| 6 | 5 | 2003 | 2 | 1 | 5 | 2 | 1956 | 47 | 3 |
| 6 | 5 | 2003 | 1 | 2 | 1 | 1 | 1939 | 64 | 4 |

As variáveis v070, v072 e v075 indicam, respectivamente, o mês, o número e o ano da entrevista.

A partir da observação de casos como os apresentados acima, sentimos a necessidade de montar um algoritmo mais avançado de emparelhamento. A idéia é eliminar ao máximo o “falso atrito” gerado na amostra decorrente de erros nas informações declaradas, tendo o cuidado, porém, de não emparelhar pessoas distintas. Este algoritmo, com sua sintaxe para Stata apresentada no apêndice,⁵ utiliza critérios não só de exatidão nas variáveis de identificação, mas também de proximidade nas respostas.

Para o emparelhamento, construiu-se uma nova variável, denominada p201, que identifica o mesmo indivíduo em todo o painel. Essa variável é igual ao número de ordem da pessoa (v201) no domicílio, caso ela seja observada pela primeira vez no painel durante a primeira entrevista em seu domicílio. Se a pessoa aparece pela primeira vez no painel durante a segunda entrevista no domicílio, p201 será igual a v201 somado 100; se ela aparecer pela primeira vez durante a terceira entrevista, p201 será igual a v201 somado 200; e assim sucessivamente. Além dessa variável, outras duas foram criadas, uma indicando se a mesma pessoa foi identificada na entrevista seguinte (denominada ‘forw’) e outra indicando se a mesma pessoa foi identificada na entrevista anterior (denominada ‘back’).

O algoritmo de emparelhamento nada mais é do que um processo de busca da mesma pessoa em uma posição anterior na base de dados, seguindo um critério de ordenação crescente em relação ao número da entrevista no domicílio e que otimize essa busca de acordo com as variáveis que identificam esta pessoa. Para construção da variável p201, em uma primeira rodada, emparelhamos os indivíduos na segunda entrevista com aqueles da primeira entrevista. Assim, o número de ordem da pessoa na primeira entrevista é atribuído à variável p201 dela na segunda entrevista. Na

5. Para rodar a sintaxe em Stata, sugerimos que o usuário remonte as bases de dados da PME por painel (variável v060), incluindo todas as informações da primeira à oitava entrevista. Assim, não seria exigido um uso excessivo de memória.

segunda rodada, emparelhamos os indivíduos na terceira entrevista com aqueles na segunda entrevista, atribuindo o valor de p201 de um ao outro. Este processo é repetido sucessivamente em um *loop* até emparelhar as pessoas na oitava entrevista com as na sétima entrevista. Ao final, para reconstituir os casos onde o indivíduo sai da amostra porém retorna meses depois, utiliza-se um outro *loop* semelhante ao anterior, porém retrospectivo. Esse *loop* verifica se as pessoas que apareceram pela primeira vez na oitava entrevista não são observadas em entrevistas anteriores à sétima; em seguida, verifica se as pessoas que apareceram pela primeira vez na sétima entrevista não são observadas em entrevistas anteriores à sexta; e assim sucessivamente.

No primeiro *loop*, os critérios de emparelhamento seguem uma hierarquia do mais ao menos rigoroso e do menos ao mais preciso. Assim, só participam do processo menos rigoroso aqueles indivíduos que não emparelharam no processo mais rigoroso. Além disso, a manipulação das variáveis 'forw' e 'back' garante que os indivíduos não sejam emparelhados com duas pessoas ao mesmo tempo.

As idéias de rigor e precisão geram, na realidade, um *trade-off*. Um critério rigoroso (com baixa precisão) é aquele em que as pessoas emparelhadas possuem uma alta probabilidade de serem as mesmas. Contudo, as pessoas não-emparelhadas por este critério não possuem necessariamente baixa probabilidade de serem as mesmas. Um exemplo de critério muito rigoroso e com baixa precisão é a data de nascimento. Por outro lado, um critério preciso (com baixo rigor) é aquele em que as pessoas não-emparelhadas possuem uma baixa probabilidade de serem as mesmas, porém as pessoas emparelhadas não possuem necessariamente alta probabilidade de serem as mesmas. Exemplos de critérios pouco rigorosos com alta precisão são sexo e escolaridade.

Por causa dessa diferença na classificação dos critérios, há uma hierarquia em sua utilização, de modo que os mais precisos e menos rigorosos só são adotados quando já existe poucas pessoas não-emparelhadas nos domicílios. Isso, conseqüentemente, aumenta a probabilidade de as pessoas emparelhadas serem as mesmas quando utilizado um critério pouco rigoroso.

O quadro 3 apresenta os critérios do emparelhamento avançado na ordem em que são aplicados. Cabe notar que, após a utilização dos critérios mais rigorosos, alguns filtros amostrais são utilizados. A justificativa para o primeiro filtro ("chefes, cônjuges e filhos com 25 anos ou mais") é que alguns critérios mais precisos só possuem certo rigor quando aplicados a pessoas adultas que pertencem ao núcleo familiar. Já o segundo filtro ("indivíduos em domicílio onde alguém já emparelhou") é aplicado para que famílias distintas, que residiram no mesmo domicílio, não sejam emparelhadas ao se adotarem os critérios menos rigorosos.

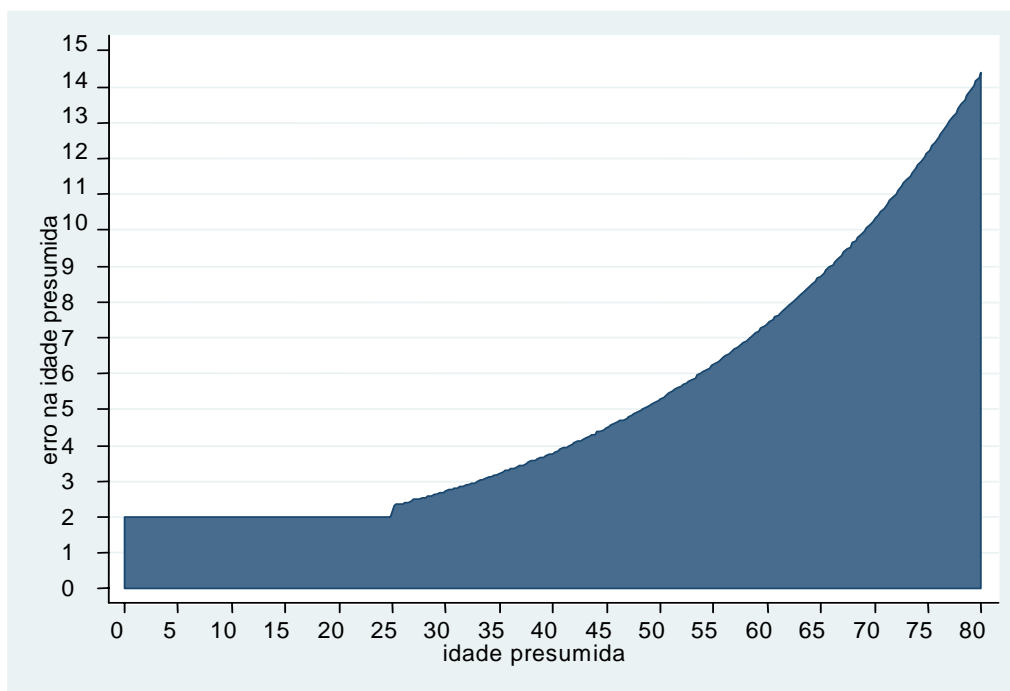
Combinação dos critérios de emparelhamento avançado

| Critérios de emparelhamento | Ordem de combinação dos critérios | | | | | | | | | | | |
|---|-----------------------------------|--|---|----------------|----------------|----------------|---|--|---|---|---|---|
| | Básico | avançado | | | | | | | | | | |
| | | Somente para chefes, cônjuges e filhos com 25 anos ou mais | | | | | | Somente para indivíduos em domicílio onde alguém já emparelhou | | | | |
| | 1 | 2 | 3 | 3 ^b | 3 ^c | 3 ^d | 4 | 4 ^e | 5 | 6 | 7 | 8 |
| Mesmo sexo | x | | x | x | x | x | x | x | x | x | x | x |
| Mesmo dia de nascimento | x | x | | | | | | | | | | |
| Até 4 dias de diferença no dia de nascimento | | | x | | x | | | | | | | |
| Mesmo mês de nascimento | x | x | | | | | | | | | | |
| Até 2 meses de diferença no mês de nascimento | | | x | x | | | | | | | | |
| Mesmo ano de nascimento | x | | | | | | | | | | | |
| Mesma idade presumida | | | | | | | x | | | | | |
| Diferença na idade presumida igual a 1 | | | | | | | | | x | | | |
| Diferença na idade presumida igual a 2 | | | | | | | | | | x | | |
| Diferença na idade presumida ^a $\leq \begin{cases} 2 & \text{se idade} < 25 \\ \exp(\text{idade}/30) & \text{se idade} \geq 25 \end{cases}$ | | | x | x | x | x | | | | | x | |
| Função acima multiplicada por dois (sse idade • 25) | | | | | | | | | | | | x |
| Mesmo número de ordem | | x | | | | | | | | | | |
| Mesma condição no domicílio | | | | | | | | x | | | | |
| Mesmo nível de escolaridade | | | | | | | | x | | | | |
| Até 1 ciclo de diferença no nível de escolaridade | | | | x | x | x | | | | | | |

^a Ver figura 1 a seguir. ^b Se dia de nascimento for não-observado. ^c Se mês de nascimento for não-observado. ^d Se dia e mês de nascimento forem não-observados. ^e Se idade presumida for não-observada.

No quadro 3, é possível notar que alguns critérios, como dia e mês de nascimento e idade, perdem rigor na medida que os emparelhamentos são executados. Por exemplo, depois de emparelhadas as pessoas que possuem a mesma data de nascimento, procuram-se aquelas que possuem datas de nascimento próximas, porém não exatamente iguais. Outro exemplo é o de erro na idade presumida em função da própria idade da pessoa. A idéia de utilizar essa função surgiu depois de constarmos que pessoas mais velhas possuem maior dificuldade em reportar a sua idade corretamente. Portanto, definimos uma função que impõe um limite máximo de diferença na idade presumida entre dois períodos. O erro aceitável para a idade presumida está demarcado na figura 1.

FIGURA 1
Área correspondente ao erro aceitável para a idade presumida



A tabela 5 mostra as taxas de atrito após a utilização do algoritmo de emparelhamento avançado, considerando todos os critérios mencionados anteriormente. Em ambas as pesquisas, é possível recuperar entre 7 p.p. e 10 p.p. da amostra nos primeiros meses de intervalo. Isso significa que a taxa de atrito cai quase pela metade ao passarmos do emparelhamento básico para o avançado. Em domicílios emparelhados, essa recuperação de indivíduos representa uma redução de mais de 60% na taxa de atrito. Nos intervalos maiores, recuperaram-se entre 5 p.p. e 6 p.p. da amostra na antiga PME e entre 15 p.p. e 25 p.p. da amostra na nova PME. Em domicílios emparelhados, isso representa reduções na taxa de atrito entre 25% e 40%, e entre 40% e 55%, respectivamente.

Após o uso do algoritmo de emparelhamento avançado, constatamos que a taxa de atrito nos intervalos maiores continua mais elevada na nova PME que na antiga. No entanto, a recuperação de observações no painel da nova pesquisa é consideravelmente mais elevada. Isso ressalta a hipótese de que o aumento relativo no desgaste de indivíduos de uma pesquisa para a outra se deve, entre outras coisas, à queda na qualidade da informação utilizada como critério para o emparelhamento.

TABELA 5

Perdas entre pares de meses como proporção do limite rotacional da PME, seguindo os critérios avançados de emparelhamento

(Em %)

| Intervalo de meses | Antiga PME | | | Nova PME | | |
|--------------------|-----------------------------|----------------------------|-------------|-----------------------------|----------------------------|-------------|
| | Percentual de indivíduos | | | Percentual de indivíduos | | |
| | Atritados (não-recuperados) | | Recuperados | Atritados (não-recuperados) | | Recuperados |
| | Total | Em domicílios emparelhados | | Total | Em domicílios emparelhados | |
| 1 | 6,0 | 2,1 | 7,3 | 6,1 | 2,7 | 6,9 |
| 2 | 8,4 | 3,3 | 7,7 | 9,3 | 4,5 | 8,6 |
| 3 | 11,2 | 4,6 | 8,1 | 12,7 | 6,1 | 10,3 |
| 9 | 15,3 | 7,6 | 5,4 | 30,2 | 20,4 | 24,7 |
| 10 | 16,6 | 8,7 | 5,7 | 30,3 | 21,1 | 24,3 |
| 11 | 18,0 | 9,7 | 5,9 | 30,8 | 21,7 | 23,9 |
| 12 | 19,6 | 10,7 | 6,2 | 31,9 | 22,2 | 23,6 |
| 13 | 21,1 | 11,9 | 6,2 | 41,2 | 23,9 | 18,4 |
| 14 | 24,7 | 14,7 | 4,6 | 45,8 | 20,2 | 18,4 |
| 15 | 26,0 | 14,7 | 5,7 | 54,6 | 20,5 | 14,3 |

Fonte: Elaboração própria a partir da PME 1998, 1999, 2003 e 2004.

Nota: Valores referentes à média das amostras presentes nos meses de janeiro a setembro de 1998, para a antiga PME, e nos meses de janeiro a dezembro de 2003, para a nova PME.

Para corrigir os problemas informacionais e facilitar a reconstituição do painel, não só da PME, mas de outras pesquisas que venham a ser realizadas neste formato, sugerimos que o IBGE adote um processo de crítica dos dados que garanta uma consistência nas informações entre as entrevistas. Por exemplo, a data de nascimento de uma pessoa entrevistada em janeiro teria de ser consistente com a data declarada em fevereiro. Além disso, seria interessante o uso de um código único por pessoa que a identifique da primeira à oitava entrevista. Com isso, seria possível perguntar, inclusive, a razão pela ausência ou inclusão da pessoa no domicílio. Por fim, acreditamos que estas sugestões não são de difícil operacionalização, principalmente porque o IBGE vem adotando o uso de *Personal Digital Assistants* (PDAs) na coleta de informações desde 2002.

4 IMPLICAÇÃO DE MUDANÇAS NO EMPARELHAMENTO SOBRE ALGUNS RESULTADOS

Se o erro na informação que impossibilita o emparelhamento de pessoas no painel ocorre de maneira aleatória, podemos dizer que a utilização do algoritmo de emparelhamento avançado possui apenas a vantagem de aumentar o tamanho da amostra, aumentando assim a eficiência em análises de painel. Caso contrário, além da redução no tamanho amostral, a utilização apenas do emparelhamento básico resultaria em análises viesadas.

Para investigar a ocorrência deste viés, as tabelas 6 e 7 apresentam matrizes de transições de posição na ocupação, mensais e anuais, respectivamente, calculadas com base em dois painéis construídos. Ambos os painéis são de adultos, entre 18 e 60 anos, que estavam inicialmente ocupados no mercado de trabalho. Contudo, o primeiro painel foi construído com base no algoritmo de emparelhamento básico, enquanto o segundo painel foi construído com base no algoritmo de emparelhamento avançado.

Nas transições mensais (tabela 6), apesar de a amostra aumentar 11% quando se utiliza o emparelhamento avançado, nenhuma diferença considerável entre os resultados é identificada. Portanto, neste caso em específico, o chamado “falso atrito” entre os meses parece ser ortogonal aos resultados de interesse.

TABELA 6
Transições mensais de posição na ocupação

| Mês anterior | Posição na ocupação (%) | | | | | | | Total |
|--------------------------------------|-------------------------|--------------|--------------|--------------|---------------|------------|----------------|-------|
| | Inativo | Desempregado | Empregado SC | Empregado CC | Conta-própria | Empregador | Não-remunerado | |
| Resultado do emparelhamento básico | | | | | | | | |
| Empregado SC | 5,46 | 3,27 | 73,20 | 10,77 | 6,10 | 0,90 | 0,31 | 100 |
| Empregado CC | 1,96 | 0,99 | 4,19 | 91,60 | 0,93 | 0,31 | 0,02 | 100 |
| Conta-própria | 5,98 | 2,09 | 7,87 | 2,70 | 76,69 | 4,25 | 0,42 | 100 |
| Empregador | 1,97 | 0,51 | 4,27 | 3,17 | 15,83 | 73,62 | 0,63 | 100 |
| Não-remunerado | 14,26 | 2,96 | 11,35 | 2,41 | 11,76 | 4,72 | 52,54 | 100 |
| Total | 3,64 | 1,73 | 21,14 | 50,47 | 17,49 | 4,96 | 0,57 | 100 |
| Número de observações | 1.700.806 | | | | | | | |
| Resultado do emparelhamento avançado | | | | | | | | |
| Empregado SC | 5,68 | 3,36 | 72,52 | 11,03 | 6,19 | 0,90 | 0,31 | 100 |
| Empregado CC | 2,12 | 1,09 | 4,40 | 91,03 | 1,01 | 0,33 | 0,02 | 100 |
| Conta-própria | 6,22 | 2,18 | 8,06 | 2,94 | 75,85 | 4,32 | 0,43 | 100 |
| Empregador | 2,10 | 0,52 | 4,42 | 3,35 | 16,33 | 72,63 | 0,65 | 100 |
| Não-remunerado | 14,52 | 3,04 | 11,43 | 2,47 | 11,99 | 4,82 | 51,73 | 100 |
| Total | 3,83 | 1,82 | 21,34 | 50,23 | 17,33 | 4,86 | 0,57 | 100 |
| Número de observações | 1.886.293 | | | | | | | |

Fonte: Elaboração própria a partir da PME de 2002-2007.

Nota: SC = sem carteira; CC = com carteira.

Nas transições anuais (tabela 7), podemos notar que as taxas de permanência na mesma posição estão significativamente sobreestimadas quando se utiliza a amostra com emparelhamento básico. Em contrapartida, as taxas de transição para a posição de empregado com ou sem carteira assinada e, principalmente, para a inatividade estão subestimadas. Além disso, a amostra derivada do emparelhamento avançado 55% maior que a amostra derivada do emparelhamento básico. Independentemente do viés, esse aumento na amostra pode significar um incremento considerável na

eficiência de estimativas, principalmente quando o pesquisador trabalhar com uma amostra muito restrita, em geral mulheres, entre 20 e 30 anos, casadas, com filhos.

TABELA 7
Transições anuais de posição na ocupação

| Mês anterior | Posição na ocupação (%) | | | | | | | Total |
|--------------------------------------|-------------------------|--------------|--------------|--------------|---------------|------------|----------------|-------|
| | Inativo | Desempregado | Empregado SC | Empregado CC | Conta-própria | Empregador | Não-remunerado | |
| Resultado do emparelhamento básico | | | | | | | | |
| Empregado SC | 10,79 | 6,45 | 52,12 | 19,46 | 9,24 | 1,55 | 0,38 | 100 |
| Empregado CC | 4,70 | 3,49 | 5,45 | 83,55 | 2,13 | 0,64 | 0,05 | 100 |
| Conta-própria | 10,64 | 2,81 | 10,62 | 5,83 | 63,61 | 5,98 | 0,50 | 100 |
| Empregador | 3,98 | 1,05 | 5,90 | 5,27 | 20,69 | 62,29 | 0,82 | 100 |
| Não-remunerado | 19,55 | 4,82 | 14,17 | 6,37 | 17,92 | 6,80 | 30,37 | 100 |
| Total | 7,30 | 3,88 | 16,88 | 48,89 | 17,14 | 5,43 | 0,48 | 100 |
| Número de observações | 422.547 | | | | | | | |
| Resultado do emparelhamento avançado | | | | | | | | |
| Empregado SC | 11,48 | 6,50 | 51,07 | 19,55 | 9,49 | 1,53 | 0,37 | 100 |
| Empregado CC | 5,15 | 3,64 | 6,18 | 81,90 | 2,40 | 0,68 | 0,06 | 100 |
| Conta-própria | 11,51 | 3,17 | 11,29 | 6,65 | 60,72 | 6,07 | 0,58 | 100 |
| Empregador | 4,68 | 1,10 | 6,77 | 6,02 | 22,09 | 58,48 | 0,86 | 100 |
| Não-remunerado | 21,79 | 4,87 | 14,09 | 6,03 | 18,20 | 7,28 | 27,74 | 100 |
| Total | 7,95 | 4,08 | 17,63 | 48,29 | 16,53 | 5,03 | 0,48 | 100 |
| Número de observações | 656.929 | | | | | | | |

Fonte: Elaboração própria a partir da PME de 2002-2007.

Nota: SC = sem carteira; CC = com carteira.

5 CONCLUSÃO

A finalidade deste artigo era levantar alguns aspectos sobre como lidar com o painel da PME, particularmente em relação ao seu desgaste. Na reconstituição do painel através dos microdados da pesquisa, mostramos que, além das diversas razões que fazem as pessoas não serem encontradas nas entrevistas seguintes, há ainda outra fonte de desgaste ou atrito. Trata-se da imprecisão nas informações declaradas pelas pessoas entrevistadas.

A PME não reporta um código que possibilite identificar com certeza a mesma pessoa em períodos distintos. Portanto, é comum utilizarmos algumas características individuais, reportadas com certa imprecisão, para identificar a mesma pessoa em duas ou mais entrevistas. De fato, quanto maior a imprecisão nas informações que servem como uma chave de reconstituição, maior é o “falso atrito” gerado, o que sobreestima o verdadeiro desgaste do painel.

De acordo com o processo de emparelhamento convencional dos microdados, a perda de pessoas no painel no intervalo de um mês é em torno de 13%, sendo

crescente com o aumento no intervalo dos meses. A partir dos nove meses de intervalo, a taxa de atrito de indivíduos varia entre 20% e 30%, na antiga PME, e chega a mais de 50% na nova pesquisa. Com o uso de um algoritmo de emparelhamento avançado, foi possível recuperar entre 7 p.p. e 10 p.p. das amostras nos primeiros meses de intervalo, o que representa uma queda de quase 50% na taxa de atrito. Nos intervalos maiores, recuperamos entre 5 p.p. e 6 p.p. da amostra na antiga PME e entre 15 p.p. e 25 p.p. da amostra na nova PME. Portanto, parte do aumento relativo no desgaste de indivíduos de uma pesquisa para a outra se deve, de fato, à queda na qualidade da informação utilizada como critério para o emparelhamento. Contudo, mesmo após o emparelhamento avançado, a taxa de atrito nos intervalos maiores continua mais elevada na nova PME que na antiga, sugerindo que há também um efeito de mudanças no desenho e na composição amostral sobre esta taxa.

Além de reduzir consideravelmente o tamanho amostral, o erro de informação, que dificulta a reconstituição do painel, pode causar também mudanças nos resultados de algumas análises, principalmente quando se trata da investigação de mudanças anuais. Portanto, um melhor emparelhamento de pessoas, que recupere os casos com imprecisão nas informações, implica não só um aumento na eficiência, mas também na consistência de investigações.

Como sugestão para facilitar a reconstituição do painel não só da PME, mas também de outras pesquisas que venham a ser realizadas neste formato, sugerimos que o IBGE adote um processo de crítica dos dados que garanta uma consistência nas informações entre as entrevistas, além do uso de um código único por pessoa que a identifique da primeira à oitava entrevista. Com isso, seria possível perguntar ainda a razão pela ausência ou inclusão da pessoa no domicílio, o que possibilitaria avanços em pesquisas relacionadas a mudanças nas composições domiciliares.

REFERÊNCIAS

IBGE. Departamento de Emprego e Rendimento. *Para compreender a PME* (um texto simplificado). 4. ed. Rio de Janeiro, 1998.

_____. *Pesquisa Mensal de Emprego*. Rio de Janeiro, 2002 (Relatório Metodológico, v. 23).

_____. Departamento de Estatísticas de População e Sociais. *Manual do entrevistador, Pesquisa Mensal de Emprego – 1980*. Rio de Janeiro, 1980.

LOPES, M. D. *Avaliação de desgaste de painéis em estudos longitudinais*: uma aplicação na Pesquisa Mensal de Emprego (PME/IBGE). Dissertação (Mestrado)–Escola Nacional de Ciências Estatísticas, IBGE, Rio de Janeiro, 2002.

APÊNDICE

Sintaxe para reconstituição do painel da PME

* Nota

/* Este algoritmo pode ser aplicado tanto à nova quanto à antiga PME. As variáveis utilizadas a seguir são as da nova pesquisa. Para utilizar o algoritmo com a antiga PME basta substituir as seguintes variáveis:

| Na nova | Na antiga |
|--|-----------------------|
| v035 = Região Metropolitana | = v010 |
| v040 = Número de controle | = v102 |
| v050 = Número de série | = v103 |
| v060 = Painel | = deve ser construída |
| v063 = Grupo rotacional | = v106 |
| v070 = Mês da pesquisa | = v105 |
| v075 = Ano da pesquisa | = deve ser construída |
| v072 = Número da pesquisa no domicílio | = deve ser construída |
| v201 = Número de ordem | = v201 |
| v203 = Sexo | = v202 |
| v204 = Dia de nascimento | = v206 |
| v214 = Mês de nascimento | = v236 |
| v224 = Ano de nascimento | = v246 |
| v234 = Idade calculada | = v256 |
| v205 = Condição no domicílio | = v203 |
| vdae1 = Anos de estudo I | = deve ser construída |

Recomenda-se rodá-lo com arquivos pequenos. Para PME, a sugestão é de um arquivo por painel (variável v060) */

* Variáveis do painel

* Variável de identificação da pessoa no painel

g p201 = v201 if v072 == 1 /* definido com base na 1a entrevista */

* Variáveis que identificam o emparelhamento

g back = . /* com uma entrevista anterior */

g forw = . /* com uma entrevista posterior */

* Emparelhamento - 1a loop

* Emparelhamento para cada par de entrevista por vez

for values i = 1/7 {

* Emparelhamento-padrão - se a data de nascimento está correta

* Ordenando cada indivíduo pelo mês de entrevista

sort v035 v040 v050 v060 v063 v203 v204 v214 v224 v075 v070 v201

* Loop para procurar a mesma pessoa em uma posição anterior

loc j = 1 /* j determina a posição anterior na base */

loc stop = 0 /* se stop=1, a loop para */

loc count = 0

while `stop' == 0 {

loc lastcount = `count'

count if p201 == . & v072 == `i'+1 /* observações não
emparelhadas */

loc count = r(N)

if `count' == `lastcount' {

 * Parar caso a loop não esteja emparelhando mais

 loc stop = 1

 }

else {

 if r(N) != 0 {

 * Captando a identificação p201 da observação anterior

 replace p201 = p201[_n - `j'] if /*

 Identificação do domicílio

 */ v035 == v035[_n - `j'] & ///

 v040 == v040[_n - `j'] & ///

 v050 == v050[_n - `j'] & ///

 v060 == v060[_n - `j'] & ///

 v063 == v063[_n - `j'] & /*

diferença entre períodos */ v072 == `i'+1 & v072[_n - `j'] == `i' /*

excluir emparelhados */ & p201 == . & forw[_n - `j'] != 1 & /*

 Características individuais

 Sexo */ v203 == v203[_n - `j'] & /*

 Dia de nascimento */ v204 == v204[_n - `j'] & /*

 Mês de nascimento */ v214 == v214[_n - `j'] & /*

 Ano de nascimento */ v224 == v224[_n - `j'] & /*

 Informação observada */ v204!=99 & v214!=99 & v224!=9999

 * Identificação de emparelhamento para quem está à frente

 replace forw = 1 if v035 == v035[_n + `j'] & ///

 v040 == v040[_n + `j'] & ///

```

v050 == v050[_n + `j'] & ///
v060 == v060[_n + `j'] & ///
v063 == v063[_n + `j'] & ///
p201 == p201[_n + `j'] & ///
v072 == `i' & v072[_n + `j']==`i'+1 ///
& forw != 1

    loc j = `j' + 1 /* passando para a próxima observação */
}
else {

    * Parar se não há observações para emparelhar
    loc stop = 1
}
}
}

* Recodificar variáveis de identificação do emparelhamento
replace back = p201 !=. if v072 == `i'+1
replace forw = 0 if forw != 1 & v072 == `i'

*****
* Emparelhamento avançado
*****

* Se sexo e ano de nascimento não estiverem corretos

* Isolando observações já emparelhadas
tempvar aux
g `aux' = (forw==1 & (v072==1 | back==1)) | (back==1 & v072==8)

* Ordenando cada indivíduo pelo mês de entrevista
sort `aux' v035 v040 v050 v060 v063 v204 v214 v201 v075 v070

* Loop para procurar a mesma pessoa em uma posição anterior
loc j = 1 /* j determina a posição anterior na base */
loc stop = 0 /* se stop=1, a loop para */
loc count = 0
while `stop' == 0 {
    loc lastcount = `count'
    count if p201 == . & v072 == `i'+1 /* observações não
                                emparelhadas */

    loc count = r(N)
    if `count' == `lastcount' {

        * Parar caso a loop não esteja emparelhando mais
        loc stop = 1
    }
}

```

```

else {
  if r(N) != 0 {

    * Captando a identificação p201 da observação anterior
    replace p201 = p201[_n - `j'] if /*
      Identificação do domicílio
      */ v035 == v035[_n - `j'] & ///
      v040 == v040[_n - `j'] & ///
      v050 == v050[_n - `j'] & ///
      v060 == v060[_n - `j'] & ///
      v063 == v063[_n - `j'] & /*
    diferença entre períodos */ v072 == `i'+1 & v072[_n - `j'] == `i' /*
    excluir emparelhados */ & p201 ==. & forw[_n - `j'] != 1 & /*
      Características individuais
    Dia de nascimento */ v204 == v204[_n - `j'] & /*
    Mês de nascimento */ v214 == v214[_n - `j'] & /*
    Mesmo número de ordem */ v201 == v201[_n - `j'] & /*
    Informação observada */ v204!=99 & v214!=99

    * Identificação de emparelhamento para quem está à frente
    replace forw = 1 if v035 == v035[_n + `j'] & ///
      v040 == v040[_n + `j'] & ///
      v050 == v050[_n + `j'] & ///
      v060 == v060[_n + `j'] & ///
      v063 == v063[_n + `j'] & ///
      p201 == p201[_n + `j'] & ///
      v072 == `i' & v072[_n + `j'] == `i'+1 ///
      & forw != 1

    loc j = `j' + 1 /* passando para a próxima observação */
  }
  else {

    * Parar se não há observações para emparelhar
    loc stop = 1
  }
}
}

*****
* Emparelhamento avançado
*****
* Somente para chefes, cônjuges e filhos adultos

tempvar ager aux

```

* Função de erro na idade presumida

```
g `ager' = cond(v234>=25 & v234<999, exp(v234/30), 2)
```

* Isolando observações já emparelhadas

```
g `aux' = (forw==1 & (v072==1 | back==1)) | (back==1 & v072==8)
```

* Ordenando cada família pelo mês de entrevista

```
sort `aux' v035 v040 v050 v060 v063 v203 v075 v070 v234 vdae1 v201
```

* Loop para procurar a mesma pessoa em uma posição anterior

```
loc j = 1
```

```
loc stop = 0
```

```
loc count = 0
```

```
while `stop' == 0 {
```

```
    loc lastcount = `count'
```

```
    count if p201==. & v072==`i'+1 & ///
```

```
(v205<=2 | (v205==3 & v234>=25 ///
```

```
& v234<999)) /* observações não emparelhadas */
```

```
    loc count = r(N)
```

```
    if `count' == `lastcount' {
```

```
        loc stop = 1
```

```
    }
```

```
    else {
```

```
        if r(N) != 0 {
```

```
            replace p201 = p201[_n - `j'] if /*
```

```
                Identificação do domicílio
```

```
                */ v035 == v035[_n - `j'] & ///
```

```
                v040 == v040[_n - `j'] & ///
```

```
                v050 == v050[_n - `j'] & ///
```

```
                v060 == v060[_n - `j'] & ///
```

```
                v063 == v063[_n - `j'] & /*
```

```
Diferença entre períodos */ v072 == `i'+1 & v072[_n - `j'] == `i' /*
```

```
Excluir emparelhados */ & p201 ==. & forw[_n - `j'] != 1 & /*
```

```
                Características individuais
```

```
Sexo */ v203 == v203[_n - `j'] & /*
```

```
Diferença na idade */ abs(v234 - v234[_n - `j'])<=`ager' & /*
```

```
Idade observada */ v234!=999 & /*
```

```
Se chefe ou cônjuge */ ((v205<=2 & v205[_n - `j']<=2) | /*
```

```
ou filho com mais de 25 */ (v234>=25 & v234[_n - `j']>=25 & ///
```

```
v205==3 & v205[_n - `j']==3)) & /*
```

```
Até 4 dias de erro na data */ ((abs(v204 - v204[_n - `j'])<=4 & /*
```

```
Até 2 meses de erro na data*/ abs(v214 - v214[_n - `j'])<=2 & /*
```

```
Informação observada */ v204!=99 & v214!=99) /*
```

```
ou */ | /*
```

```
1 ciclo de erro na educação*/ (abs(vdae1 - vdae1[_n - `j'])<=1 /*
```

```
e */ & /*
```

```
Até 2 meses de erro na data*/ ((abs(v214 - v214[_n - `j'])<=2 & /*
```



```

Informação observada */    v214!=99 & /*
Informação não observada */    (v204==99 | v204[_n-`j']==99)) /*
ou */                        | /*
até 4 dias de erro na data */ (abs(v204 - v204[_n - `j'])<=4 & /*
Informação observada */    v204!=99 & /*
Informação não-observada */    (v214==99 | v214[_n - `j']==99)) /*
ou */                        | /*
informações não-observadas */ ((v204==99 | v204[_n - `j']==99) & ///
                                (v214==99 | v214[_n - `j']==99))))

    replace forw = 1 if v035 == v035[_n + `j'] & ///
                    v040 == v040[_n + `j'] & ///
                    v050 == v050[_n + `j'] & ///
                    v060 == v060[_n + `j'] & ///
                    v063 == v063[_n + `j'] & ///
                    p201 == p201[_n + `j'] & ///
                    v072 == `i' & v072[_n + `j']==`i'+1 ///
                    & forw != 1
    loc j = `j' + 1
    }
    else {
        loc stop = 1
    }
}
}
replace back = p201 !=. if v072 == `i'+1
replace forw = 0 if forw != 1 & v072 == `i'

*****
* Emparelhamento avançado
*****

* Somente em domicílio onde alguém já emparelhou

* Quantas pessoas emparelharam no domicílio
tempvar dom
bys v075 v070 v035 v040 v050 v060 v063: egen `dom' = sum(back)

* Loop com os critérios de emparelhamento
foreach w in /*mesma idade*/ "0" /*erro na idade = 1*/ "1" /*
erro na idade = 2*/ "2" /*erro na idade = f(idade)*/ "`ager'" /*
2xf(idade)*/ "2*`ager' & v234>=25" {

    * Isolando observações já emparelhadas
tempvar aux
g `aux' = (forw==1 & (v072==1 | back==1)) | ///
(back==1 & v072==8) | (`dom'==0 & v072==`i'+1)

```

```

sort `aux' v035 v040 v050 v060 v063 v203 v075 v070 v234 ///
vdae1 v201
loc j = 1
loc stop = 0
loc count = 0
while `stop' == 0 {
    loc lastcount = `count'
    count if p201 == . & v072 == `i'+1 & `dom'>0 & `dom'!=.
    loc count = r(N)
    if `count' == `lastcount' {
        loc stop = 1
    }
    else {
        if r(N) != 0 {
            replace p201 = p201[_n - `j'] if /*
                Identificação do domicílio
                */ v035 == v035[_n - `j'] & ///
                v040 == v040[_n - `j'] & ///
                v050 == v050[_n - `j'] & ///
                v060 == v060[_n - `j'] & ///
                v063 == v063[_n - `j'] & /*
Diferença entre períodos */    v072 == `i'+1 & v072[_n-`j'] == `i' /*
excluir emparelhados */        & p201 ==. & forw[_n - `j'] != 1 & /*
há emparelhados no domicílio*/ `dom' > 0 & `dom'!=. & /*
                Características individuais
Sexo */                v203 == v203[_n - `j'] & /*
Critérios mudam com a loop */ ((abs(v234-v234[_n - `j'])<=`w' & /*
se a idade é observada */    v234!=999) /*
caso contrário */        | /*
Mesma escolaridade */        (vdae1==vdae1[_n - `j'] & /*
Mesma condição no domicílio */ v205==v205[_n - `j'] & /*
Idade não observada */        (v234==999 | v234[_n - `j']==999)))

            replace forw = 1 if v035 == v035[_n + `j'] & ///
                v040 == v040[_n + `j'] & ///
                v050 == v050[_n + `j'] & ///
                v060 == v060[_n + `j'] & ///
                v063 == v063[_n + `j'] & ///
                p201 == p201[_n + `j'] & ///
                v072 == `i' & v072[_n+`j']==`i'+1 ///
                & forw != 1

            loc j = `j' + 1
        }
    }
    else {
        loc stop = 1
    }
}
}

```

```

    }
  }
  replace back = p201 !=. if v072 == `i'+1
  replace forw = 0 if forw != 1 & v072 == `i'

  * Identificação para quem estava ausente na última entrevista
  replace p201 = `i'00 + v201 if p201 == . & v072 == `i'+1
}

*****
* Recuperar quem saiu e retornou para o painel - 2a loop
*****

* Variável temporária identificando o emparelhamento à frente
tempvar fill
g `fill' = forw

* Loop retrospectivo por entrevista
foreach i in 7 6 5 4 3 2 1 {
  tempvar ncode1 ncode2 aux max ager

  * Função de erro na idade presumida
  g `ager' = cond(v234>=25 & v234<999, exp(v234/30), 2)

  * Variável que preserva o antigo número
  bys v035 v040 v050 v060 v063 p201: g `ncode1' = p201

  * Isolando observações emparelhadas
  g `aux' = ((`fill'==1 & (v072==1 | back==1)) | (back==1 & v072==8))

  * Variável identificando a última entrevista
  bys v035 v040 v050 v060 v063 p201: egen `max' = max(v072)

  sort `aux' v035 v040 v050 v060 v063 v203 v072 v201 p201
  loc j = 1
  loc stop = 0
  loc count = 0
  while `stop' == 0 {
    loc lastcount = `count'
    count if p201>`i'00 & p201<`i'99 & back==0
    loc count = r(N)
    if `count' == `lastcount' {
      loc stop = 1
    }
  }
  else {
    if r(N) != 0 {
      replace p201 = p201[_n - `j'] if /*

```

```

Identificação do domicílio
*/ v035 == v035[_n - `j'] & ///
v040 == v040[_n - `j'] & ///
v050 == v050[_n - `j'] & ///
v060 == v060[_n - `j'] & ///
v063 == v063[_n - `j'] & /*

Quem entrou na entrevista */ p201>`i'00 & p201<`i'99 & /*
Não emparelhado */ back==0 & `fill'[_n - `j']!=1 & /*
Uma entrev. de diferença */ `max'[_n - `j']<`i' & ///
p201[_n - `j']<`i'00-100 & /*

Sexo */ v203 == v203[_n - `j'] & /*
Diferença na idade */ ((abs(v234 - v234[_n - `j'])<=`ager' & /*
Idade observada */ v234!=999 & /*
Até 4 dias de erro na data */ ((abs(v204 - v204[_n - `j'])<=4 & /*
Até 2 meses de erro na data*/ abs(v214 - v214[_n - `j'])<=2 & /*
informação observada */ v204!=99 & v214!=99) /*
ou */ | /*
1 ciclo de erro na educação*/ (abs(vdae1 - vdae1[_n - `j'])<=1 /*
e */ & /*
Até 2 meses de erro na data*/ ((abs(v214 - v214[_n - `j'])<=2 & /*
Informação observada */ v214!=99 & /*
Informação não-observada */ (v204==99 | v204[_n - `j']==99) /*
ou */ | /*
até 4 dias de erro na data */ (abs(v204 - v204[_n - `j'])<=4 & /*
Informação observada */ v204!=99 & /*
Informação não-observada */ (v214==99 | v214[_n - `j']==99) /*
ou */ | /*
nada é observado */ ((v204==99 | v204[_n - `j']==99) & ///
(v214==99 | v214[_n - `j']==99)))) /*
ou */ | /*
mesma escolaridade */ (vdae1==vdae1[_n - `j'] & /*
e número de ordem */ v205==v205[_n - `j'] /*
Se idade não é observada */ & (v234==999 | v234[_n - `j']==999)))

```

```

* Identificação de emparelhamento para quem está à frente
replace `fill' = 1 if v035 == v035[_n + `j'] & ///
v040 == v040[_n + `j'] & ///
v050 == v050[_n + `j'] & ///
v060 == v060[_n + `j'] & ///
v063 == v063[_n + `j'] & ///
p201 == p201[_n + `j'] & ///
`fill' == 0 & `max'<`i' & ///
(v072[_n + `j'] - v072)>=2

loc j = `j' + 1
}
else {
loc stop = 1

```

```
}  
}  
}
```

```
* Igualando o número de quem era igual  
bys v035 v040 v050 v060 v063 `ncode1': egen `ncode2' = min(p201)  
replace p201 = `ncode2'  
}
```

```
*****
```

```
* Fim do arquivo
```

```
*****
```

EDITORIAL

Coordenação

Iranilde Rego

Supervisão

Andrea Bossle de Abreu

Revisão

Lucia Duarte Moreira

Alejandro Sainz de Vicuña

Eliezer Moreira

Elisabete de Carvalho Soares

Míriam Nunes da Fonseca

Editoração

Roberto das Chagas Campos

Aeromilson Mesquita

Camila Guimarães Simas

Camila Oliveira de Souza

Carlos Henrique Santos Vianna

Brasília

SBS – Quadra 1 – Bloco J – Ed. BNDES,

9ª andar – 70076-900 – Brasília – DF

Fone: (61) 3315-5090

Fax: (61) 3315-5314

Correio eletrônico: editbsb@ipea.gov.br

Rio de Janeiro

Av. Nilo Peçanha, 50/609

20044-900 – Rio de Janeiro – RJ

Fone: (21) 3515-8522 – 3515-8426

Fax (21) 3515-8585

Correio eletrônico: editrj@ipea.gov.br

Tiragem: 135 exemplares