

# Mercado de Trabalho

conjuntura e análise

ANO 27 | abril de 2021

71

**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**

Carlos von Doellinger

**Diretor de Desenvolvimento Institucional**

Manoel Rodrigues Junior

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

Flávia de Holanda Schmidt

**Diretor de Estudos e Políticas  
Macroeconômicas**

José Ronaldo de Castro Souza Júnior

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

Nílo Luiz Saccaro Júnior

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

André Tortato Rauen

**Diretora de Estudos e Políticas Sociais**

Lenita Maria Turchi

**Diretor de Estudos e Relações Econômicas  
e Políticas Internacionais**

Ivan Tiago Machado Oliveira

**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**

Carlos Henrique Leite Corseuil

**Membros**

Felipe Mendonça Russo

Lauro Ramos

Sandro Pereira Silva

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 2021

---

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: Boletim Mercado de Trabalho (BMT)

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

---

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.

# ESTIMAÇÃO DE FLUXOS OCUPACIONAIS BRUTOS COM A PNAD CONTÍNUA: COMPENSANDO A NÃO RESPOSTA E CONSIDERANDO O PLANO AMOSTRAL<sup>1</sup>

Guilherme A. P. Jacob<sup>2</sup>  
Pedro Luis do Nascimento Silva<sup>3</sup>

## 1 INTRODUÇÃO

Análises da evolução de indicadores sobre o mercado de trabalho têm papel de destaque nas análises estruturais e de conjuntura sobre o desempenho da economia, ademais de fornecerem insumo essencial para tomada de decisões tanto na esfera pública como na privada. Em muitos países, tais indicadores são elaborados com base em pesquisas domiciliares sobre a força de trabalho. No Brasil esse papel cabe à Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) (IBGE, 2020).

Entre os indicadores produzidos pela PNAD Contínua, as estimativas dos totais de pessoas classificadas por situação ocupacional a cada trimestre, bem como suas variações têm papel de destaque. Nesta nota, consideramos que um indivíduo membro da população em idade ativa (quinze anos e mais) só pode ser classificado em uma de três categorias segundo sua situação ocupacional: 1 – ocupado; 2 – desocupado; ou 3 – fora da força de trabalho (IBGE, 2020).

O total estimado de pessoas desocupadas é um exemplo de estimativa de estoque ao estimar o contingente de pessoas na situação ocupacional desocupado num dado momento. A diferença dos totais estimados de pessoas desocupadas em dois trimestres consecutivos é exemplo de uma estimativa de fluxo, em particular, de um fluxo líquido. Análises de fluxos líquidos, no entanto, são incompletas por não revelarem as entradas e saídas de pessoas da categoria desocupado da situação ocupacional. A título de exemplo, um fluxo líquido pode ser nulo quando 1 milhão de pessoas passaram a desocupadas no trimestre atual, e outro 1 milhão de pessoas desocupadas passaram a ocupadas no trimestre atual.

Uma descrição mais completa da dinâmica da situação ocupacional é obtida estimando os fluxos brutos, que representam os contingentes de pessoas classificadas por situação ocupacional em dois períodos, como um trimestre base e seu trimestre posterior. A título de exemplo, ao estimar o número de pessoas desocupadas no trimestre base que permaneceram desocupadas, ou tornaram-se ocupadas, ou saíram da força de trabalho no trimestre seguinte, teríamos estimativas dos fluxos brutos tendo como categoria base a situação ocupacional desocupado.

1. DOI: <http://dx.doi.org/10.38116/bmt71/nta1>

2. Doutorando no Programa de Pós-Graduação em População, Território e Estatísticas Públicas na Escola Nacional de Ciências Estatísticas do Instituto Brasileiro de Geografia e Estatística (ENCE/IBGE). *E-mail*: <guilhermejacob91@gmail.com>.

3. Secretário da Sociedade para o Desenvolvimento da Pesquisa Científica (SCIENCE), e pesquisador titular colaborador voluntário na ENCE/IBGE.

Nesta nota tratamos do problema da estimação de fluxos brutos da situação ocupacional usando os dados da PNAD Contínua. Os fluxos brutos entre dois trimestres consecutivos são os contingentes de pessoas que tinham situação ocupacional  $i$  no trimestre base, e passaram a ter situação ocupacional  $j$  no trimestre seguinte, com  $i, j = 1, 2$  ou  $3$ . Quando  $i=j$  dizemos que houve permanência na situação ocupacional  $i$ , e quando  $i \neq j$  dizemos que houve uma transição de  $i$  para  $j$  ao longo do trimestre. Nesta análise não vamos considerar movimentações que possam ter ocorrido dentro do trimestre, considerando apenas os dados referentes à situação ocupacional observada/registrada nas semanas de referência em que o indivíduo foi entrevistado na pesquisa.

Fluxos líquidos também podem ser definidos como diferenças entre a soma dos fluxos brutos de entrada menos a soma dos fluxos brutos de saída de determinada categoria. Sendo assim, a estimação dos fluxos brutos é uma descrição mais rica e completa da dinâmica, já que permite a subsequente estimação dos fluxos líquidos.

A confusão entre fluxos líquidos e brutos é comum. Não é difícil encontrar artigos com afirmações do tipo “1 milhão de pessoas entraram (ou saíram) na pobreza”, baseadas (incorretamente) na diferença de estoques de pessoas pobres entre dois períodos de comparação. Por este número se tratar de um fluxo líquido, o número de pessoas que entraram na pobreza pode ser muito maior, mas é reduzido pelo total de pessoas que saíram da pobreza no mesmo período.

Com pesquisas amostrais repetidas que usam painéis – sejam eles fixos ou rotativos –, é possível obter informações sobre a situação ocupacional dos indivíduos em períodos consecutivos. Sob amostragem aleatória simples e resposta completa, estimativas não enviesadas dos fluxos brutos poderiam ser facilmente obtidas. No entanto, pesquisas de força de trabalho geralmente utilizam planos amostrais complexos e estão sempre sujeitas à ocorrência de não resposta em um ou até em ambos os períodos de interesse.

Embora existam soluções teóricas para lidar com estes problemas (amostragem complexa e não resposta), sua implementação não é trivial e não estava disponível para uso geral. Este artigo relata uma aplicação do pacote *surf*, desenvolvido para estimar fluxos brutos considerando não resposta e planos amostrais complexos, para estimar fluxos brutos de situação ocupacional a partir da PNAD Contínua do IBGE. Essa aplicação permitiu revelar o viés causado por ignorar a não resposta diferencial e/ou o plano amostral complexo nas análises de fluxos brutos e líquidos de situação ocupacional no Brasil.

A seção 2 apresenta uma breve descrição da abordagem proposta por Stasny (1987) para estimação de fluxos brutos com não resposta sob amostragem aleatória simples, e a extensão de Gutiérrez (2014) dessa abordagem para lidar com pesquisas que empreguem planos amostrais complexos. A seção 3 apresenta uma aplicação da abordagem estendida proposta por Gutiérrez (2014) para estimar fluxos brutos de situação ocupacional com a PNAD Contínua. A seção 4 discute a aplicabilidade do método e sugere possíveis extensões.

## 2 MODELO E MÉTODO DE AJUSTE

Seja  $U$  uma população finita com  $N$  indivíduos que podem ser classificados nas categorias  $1, \dots, G$  nos períodos 1 e 2. Sob resposta completa, os fluxos brutos podem ser obtidos por meio de contagens simples das frequências das categorias observadas para os indivíduos nos períodos 1 e 2. Seja  $X_{ij}$  o fluxo bruto da categoria  $i$  para a categoria  $j$  – isto é, o número de indivíduos classificados na categoria  $i$  no período 1 e na categoria  $j$  no período 2 –, o símbolo “+” indica a soma sobre o conjunto de categorias naquela ocasião. Desta forma, os estoques de determinada categoria  $h$  nos períodos 1 e 2 podem ser representados por  $X_{h+} = \sum_j X_{hj}$  e  $X_{+h} = \sum_i X_{ih}$ , respectivamente. Essas quantidades são dispostas na tabela 1.

TABELA 1  
Estoques iniciais ( $X_{i+}$ ) e finais ( $X_{+j}$ ), e fluxos brutos ( $X_{ij}$ ) na população  $U$

Categoria no período 1	Estoques no período 1	Categoria no período 2			
		1	2	...	G
1	$X_{1+}$	$X_{11}$	$X_{12}$	...	$X_{1G}$
2	$X_{2+}$	$X_{21}$	$X_{22}$	...	$X_{2G}$
⋮	⋮	⋮	⋮		⋮
G	$X_{G+}$	$X_{G1}$	$X_{G2}$	...	$X_{GG}$
Estoques no período 2	$N$	$X_{+1}$	$X_{+2}$	...	$X_{+G}$

Fonte: adaptado de Gutiérrez (2014, p. 34).

O fluxo líquido na categoria  $h$  é representado por  $\Delta_h$  e é definido pela diferença entre os totais de pessoas que entraram na categoria  $h$  e aqueles que saíram da categoria  $h$  entre os períodos 1 e 2; isto é,  $\Delta_h = \sum_{i \neq h} X_{ih} - \sum_{j \neq h} X_{hj}$ . Como as quantidades de pessoas que permanecem na mesma categoria  $X_{hh}$  se compensam nos estoques  $X_{+h}$  e  $X_{h+}$ , este fluxo líquido também pode ser calculado usando a diferença entre os estoques final e inicial na categoria  $h$ , isto é,  $\Delta_h = X_{+h} - X_{h+}$ .

Porém, em decorrência da não resposta que se manifesta em praticamente todas as pesquisas, na prática os fluxos brutos  $X_{ij}$  não seriam observáveis. Apresentadas na tabela 2, as contagens populacionais observáveis seriam: i)  $N_{ij}$ , a contagem de indivíduos que responderam nos dois períodos e foram classificados na categoria  $i$  no período 1 e na categoria  $j$  no período 2; ii)  $R_i$ , a contagem de indivíduos que responderam no período 1 e foram classificados na categoria  $i$ , mas não responderam no período 2; iii)  $C_j$ , a contagem de indivíduos que responderam no período 2 e foram classificados na categoria  $j$ , mas não responderam no período 1; e iv)  $M$ , a contagem de indivíduos que não responderam nos períodos 1 e 2.

TABELA 2  
Contagens observáveis na população  $U$  sob não resposta

Categoria no período 1	Categoria no período 2				Complemento linhas
	1	2	...	$G$	
1	$N_{11}$	$N_{12}$	...	$N_{1G}$	$R_1$
2	$N_{21}$	$N_{22}$	...	$N_{2G}$	$R_2$
⋮	⋮	⋮		⋮	⋮
$G$	$N_{G1}$	$N_{G2}$	...	$N_{GG}$	$R_G$
Complemento colunas	$C_1$	$C_2$	...	$C_G$	$M$

Fonte: adaptado de Gutiérrez (2014, p. 20).

Stasny (1987) desenvolveu um modelo de cadeia de Markov combinando a descrição do processo gerador das contagens não observáveis (tabela 1) e o mecanismo de não resposta, relacionando assim as contagens observáveis (tabela 2) e não observáveis (tabela 1), de modo a permitir inferência adequada sobre as últimas. Nesse modelo, as contagens observadas são vistas como resultados de um processo de dois estágios: i) o primeiro estágio segue uma cadeia de Markov – isto é, um processo modelado pela distribuição inicial da variável de interesse e probabilidades de transição –, resultando nas contagens populacionais não observáveis (ou seja, sob resposta completa); e ii) um mecanismo probabilístico de não resposta, no qual a não resposta nos períodos 1 e/ou 2 varia de acordo com o fluxo bruto  $ij$ .

Os parâmetros do modelo podem ser separados em dois conjuntos. O primeiro conjunto descreve o processo Markoviano não observável de mudanças entre as categorias: o vetor  $\boldsymbol{\eta} = [\eta_i]$  descreve a distribuição de frequência dos indivíduos nas categorias no período 1; já a matriz  $\mathbf{P} = [p_{ij}]$  descreve as probabilidades de transitar da categoria  $i$  no período 1 para a categoria  $j$  no período 2. Esses parâmetros são necessários para obter a matriz de fluxos brutos esperados  $\boldsymbol{\mu} = [\mu_{ij}] = [N\eta_i p_{ij}]$ .

O segundo conjunto contém os parâmetros que descrevem o mecanismo de não resposta: i) o vetor  $\boldsymbol{\psi} = [\psi_i]$  contém as probabilidades de resposta no período 1, que podem variar com base na categoria de classificação do indivíduo no período 1; ii) a matriz  $\boldsymbol{\rho} = [\rho_{ij}]$  contém as probabilidades condicionais de que um indivíduo responda no período 2 dado que respondeu no período 1, quando teria experimentado o fluxo bruto  $ij$ ; e iii) a matriz  $\boldsymbol{\tau} = [\tau_{ij}]$  contém as probabilidades condicionais de que um indivíduo não responda no período 2 dado que não respondeu no período 1, quando teria experimentado fluxo bruto  $ij$ . A tabela 3 apresenta as proporções esperadas segundo o modelo geral de Stasny (1987) para os dados observáveis.

TABELA 3  
Proporções esperadas sob o modelo geral de Stasny (1987)

Categoria no período 1	Categoria no período 2					Complemento linhas
	1	...	$j$	...	$G$	
1	$\psi_i \rho_{ij} \eta_i p_{ij}$					$\sum_{j=1}^G \psi_i (1 - \rho_{ij}) \eta_i p_{ij}$
⋮						
$i$						
⋮						
$G$	$\sum_{i=1}^G (1 - \psi_i) (1 - \tau_{ij}) \eta_i p_{ij}$					$\sum_{i=1}^G \sum_{j=1}^G (1 - \psi_i) \tau_{ij} \eta_i p_{ij}$
Complemento colunas						

Fonte: adaptado de Gutiérrez (2014, p. 24).

Mesmo com as restrições naturais de soma unitária para vetores de proporções como  $\boldsymbol{\eta}$  e  $\boldsymbol{\psi}$ , e outras similares aplicáveis às matrizes  $\mathbf{P}$ ,  $\boldsymbol{\rho}$  e  $\boldsymbol{\tau}$ , o número de parâmetros do modelo geral excede o número de contagens observáveis disponíveis para ajuste do modelo. Para reduzir o número de parâmetros a serem estimados, Stasny (1987) propôs quatro variantes simplificadas do modelo geral para o processo de resposta/não resposta:

- Modelo A:  $\boldsymbol{\psi} = [\psi]$ ,  $\boldsymbol{\rho} = [\rho]$ ,  $\boldsymbol{\tau} = [\tau]$  – 3 parâmetros;
- Modelo B:  $\boldsymbol{\psi} = [\psi_i]$ ,  $\boldsymbol{\rho} = [\rho]$ ,  $\boldsymbol{\tau} = [\tau]$  –  $G+2$  parâmetros;
- Modelo C:  $\boldsymbol{\psi} = [\psi]$ ,  $\boldsymbol{\rho} = [\rho_i]$ ,  $\boldsymbol{\tau} = [\tau_i]$  –  $2G+1$  parâmetros; e
- Modelo D:  $\boldsymbol{\psi} = [\psi]$ ,  $\boldsymbol{\rho} = [\rho_j]$ ,  $\boldsymbol{\tau} = [\tau_j]$  –  $2G+1$  parâmetros.

Cada variante representa um conjunto de hipóteses sobre o mecanismo de não resposta. Sob o modelo A, as probabilidades de resposta inicial e de continuar respondendo (ou não respondendo) não variam de acordo com os fluxos. Sob o modelo B, as probabilidades de resposta inicial ( $\boldsymbol{\psi}$ ) variam de acordo com a categoria no período 1, mas as probabilidades de continuar respondendo ( $\boldsymbol{\rho}$ ) e de continuar não respondendo ( $\boldsymbol{\tau}$ ) não variam com os fluxos. Sob o modelo C, a probabilidade de resposta inicial não varia entre as categorias, mas as probabilidades de continuar respondendo ( $\boldsymbol{\rho}$ ) e não respondendo ( $\boldsymbol{\tau}$ ) variam de acordo com a categoria no período 1.

Analogamente, sob o modelo D, a probabilidade de resposta inicial não varia entre as categorias, mas as probabilidades de continuar respondendo ( $\boldsymbol{\rho}$ ) e não respondendo ( $\boldsymbol{\tau}$ ) variam de acordo com a categoria no período 2. Os modelos A e B possuem graus de liberdade suficientes para testar a bondade do ajuste usando o teste  $\chi^2$  de Pearson, mas os modelos C e D não possuem graus de liberdade suficientes para aplicação do mesmo teste.

O método proposto por Stasny (1987) para estimação dos parâmetros dos modelos foi desenvolvido sob a hipótese de amostragem aleatória simples. No entanto, como mencionado anteriormente, a maioria das pesquisas sobre força de trabalho utilizam algum plano amostral complexo. De fato, Stasny (1987, p. 364) comentou que a aplicação usando a *Labor Force Survey* realizada naquele artigo tinha finalidade apenas ilustrativa, já que essa pesquisa utilizava plano amostral com múltiplos estágios de conglomeração.

Gutiérrez (2014) estendeu a abordagem de Stasny (1987) para lidar com planos amostrais complexos utilizando o método da Máxima Pseudo-Verossimilhança (Binder, 1983; Molina e Skinner, 1992), permitindo assim obter estimativas consistentes para os parâmetros populacionais e, sob a especificação correta do modelo, também para os parâmetros do modelo de superpopulação (Särndal, Swensson e Wretman, 1992). Gutiérrez (2014) apresentou uma aplicação desta abordagem desenvolvida usando os dados da pesquisa mensal de emprego (PME), com bons resultados. Um dos principais achados foi que o modelo A não se ajustava bem aos dados da PME, indicando que ignorar a não resposta resultava em estimativas viesadas dos fluxos brutos.

Embora Gutiérrez (2014) tenha apresentado evidências a favor da abordagem estendida que propôs, os algoritmos necessários para estimar os parâmetros e as variâncias das estimativas considerando planos amostrais complexos não são triviais. Usando o pacote *survey* (Lumley, 2004) da linguagem *R* como base, Jacob (2021) desenvolveu o pacote *surf* para simplificar o processo de estimação de fluxos brutos com dados de pesquisas amostrais complexas e sujeitos a não resposta que pode ou não ser diferencial, contemplando o ajuste das quatro variantes do modelo geral consideradas por Stasny (1987).

### 3 ESTIMAÇÃO DOS FLUXOS

#### 3.1 Materiais

A PNAD Contínua é a principal pesquisa socioeconômica por amostragem domiciliar em andamento no Brasil. Iniciada em 2012, emprega plano amostral complexo e usa painéis rotativos trimestrais no esquema “1-2(5)”: um domicílio é entrevistado em um determinado mês e passa dois meses fora da amostra, repetindo esse padrão por cinco trimestres – ou seja, deixando a amostra definitivamente ao fim de cinco entrevistas. Esse padrão de rotação gera uma sobreposição esperada de 80% da amostra em cada par de trimestres consecutivos. Cada participação de um domicílio na pesquisa é denominada aqui de “visita”.

Domicílios possuem identificador longitudinal, mas os moradores não. Para criar um identificador longitudinal de moradores, Jacob (2021) desenvolveu um algoritmo que associa moradores ao longo das cinco visitas, usando o identificador longitudinal do domicílio, mais o sexo, a idade e a data de nascimento do morador. Esse processo de pareamento ao longo de cinco visitas impede a análise de períodos mais recentes, já que apenas um dos cinco grupos de rotação terá realizado cinco visitas no último trimestre. Para calcular estimativas mais precisas a cada trimestre, foram utilizados os dados dos domicílios que estavam nas visitas de um a quatro no trimestre inicial. Portanto, as estimativas correspondem ao período entre o 1º trimestre de 2013 e o 4º trimestre 2018. Na seção de resultados, apresentamos somente estimativas considerando o 4º trimestre de 2018 como base, por razões de espaço. O conjunto completo de resultados obtidos está disponível em Jacob (2021).



Como alguns domicílios são perdidos ao longo das visitas, utilizou-se um modelo logístico para propensão de resposta ao nível de domicílios (Lepkowski, Kalton e Kasprzyk, 1989; Teixeira *et al.*, 2019) usando características da pessoa de referência e o tamanho do domicílio como variáveis preditoras, além de calibrar os pesos ajustados por esse modelo nos totais de pessoas usados também na pós-estratificação usual da PNAD Contínua. Os algoritmos em linguagem *R* para preparação dos dados, pareamento e correção de não resposta dos domicílios podem ser encontrados em Jacob (2021).

Para analisar os efeitos da modelagem para a não resposta e do plano amostral sobre o viés das estimativas, consideramos três cenários, descritos a seguir.

- Cenário 1: estimação do modelo A que supõe não resposta completamente aleatória ajustado às contagens amostrais dos indivíduos que responderam em ambos os períodos – corresponde a ignorar completamente a não resposta na análise.
- Cenário 2: estimação do modelo A que supõe não resposta completamente aleatória ajustado às contagens populacionais estimadas dos indivíduos que responderam em ambos os períodos, usando os pesos básicos do plano amostral, sem correção pela perda de domicílios e/ou pessoas.
- Cenário 3: estimação do modelo C ajustado às contagens populacionais estimadas observáveis dos indivíduos (respondentes ou não), com os pesos corrigidos pela perda de domicílios.

### 3.2 Resultados

A tabela 4 compara as estimativas das distribuições iniciais e finais do período entre o 4º trimestre de 2018 e o 1º trimestre de 2019 sob os três cenários. As estimativas são similares, com diferenças entre as distribuições iniciais e finais (fluxos líquidos) praticamente idênticas, independentemente do cenário.

Entretanto, os perfis das distribuições não são idênticos, com as proporções de ocupados e desocupados maiores nos cenários 2 e 3 que no cenário 1, e a diferença sendo compensada com a redução das porcentagens de pessoas fora da força de trabalho. A maior diferença decorre do uso dos pesos do desenho, que são usados na estimação dos cenários 2 e 3. As diferenças entre os cenários 2 e 3 decorrem da correção pela perda de domicílios sobre os pesos e da correção da não resposta de indivíduos no par de trimestres feita com base no modelo C.

TABELA 4

#### Estimativas das distribuições iniciais, finais e variação das situações ocupacionais na população, por cenário (4º trim. 2018-1º trim. 2019)

(Em %)

Situação ocupacional	Cenário 1			Cenário 2			Cenário 3		
	Inicial	Final	Variação	Inicial	Final	Variação	Inicial	Final	Variação
O	52,3	51,5	-0,8	54,5	53,7	-0,8	55,0	54,1	-0,8
D	6,2	7,0	0,8	6,9	7,8	0,9	7,1	7,9	0,7
F	41,5	41,5	0,0	38,6	38,5	0,0	37,9	38,0	0,1

Elaboração dos autores.

Obs.: O para ocupados; D para desocupados; e F para fora da força de trabalho.

O efeito dos pesos básicos também é o maior diferencial nas estimativas das probabilidades de transição. Isso é evidenciado pelo aumento em 3,6 pontos percentuais (p.p.) na probabilidade de permanecer desocupado na comparação entre os cenários 1 e 2.

TABELA 5

**Estimativas das probabilidades de transição entre as situações ocupacionais, por cenário (4º trim. 2018-1º trim. 2019)**

(Em %)

Situação inicial	Situação final								
	Cenário 1			Cenário 2			Cenário 3		
	O	D	F	O	D	F	O	D	F
O	87,9	3,9	8,2	89,1	4,1	6,8	89,3	4,0	6,7
D	27,6	44,8	27,6	26,9	48,5	24,6	26,9	48,7	24,4
F	9,3	5,2	85,5	8,4	5,7	85,9	8,3	5,7	86,0

Elaboração dos autores.

Obs.: O para ocupados; D para desocupados; e F para fora da força de trabalho.

Embora as estimativas das probabilidades iniciais e de transição sejam próximas, as estimativas da matriz de fluxo bruto diferem bastante entre os cenários, como mostra a tabela 6. Essas diferenças decorrem das estimativas do tamanho da população em idade ativa e, em menor grau, da interação entre as pequenas diferenças das estimativas de probabilidades iniciais e de transição. Novamente, as diferenças são evidenciadas pelo total de pessoas que permaneceram desocupadas: a troca de pesos constantes (cenário 1) para os pesos do plano amostral (cenário 2) aumentam essa contagem em quase 1 milhão de pessoas. As estimativas dos estoques de inicial e final de pessoas nas categorias também são bastante diferentes, com o contingente de desocupados aumentando cerca de 1,5 milhão de pessoas entre os cenários 1 e 3.

TABELA 6

**Estimativas de fluxos brutos entre as situações ocupacionais, por cenário (4º trim. 2018-1º trim. 2019)**

(Em milhares)

Situação inicial	Situação final											
	Cenário 1				Cenário 2				Cenário 3			
	Estoque inicial	O	D	F	Estoque inicial	O	D	F	Estoque inicial	O	D	F
O	88.248	77.526	3.446	7.275	92.778	82.709	3.775	6.294	93.107	83.114	3.766	6.226
D	10.536	2.905	4.721	2.910	11.798	3.175	5.722	2.901	12.055	3.246	5.869	2.940
F	70.033	6.531	3.654	59.848	65.622	5.482	3.751	56.389	64.218	5.335	3.674	55.209
<b>Estoque final</b>	<b>86.962</b>	<b>11.821</b>	<b>70.033</b>		<b>91.367</b>	<b>13.248</b>	<b>65.584</b>		<b>91.695</b>	<b>13.309</b>	<b>64.375</b>	

Elaboração dos autores.

Obs.: O para ocupados; D para desocupados; e F para fora da força de trabalho.

O cenário 3 também produz informações importantes sobre o mecanismo de não resposta. Segundo o modelo C ajustado nesse cenário, a probabilidade de resposta no trimestre inicial é de 89,6% independente das classificações nesse período. Já as probabilidades condicionais de responder no 1º trimestre de 2019 dado que respondeu no 4º trimestre 2018 são 92,4%, 90,6% e 93,6% para ocupados, desocupados e fora da força no trimestre inicial, respectivamente, variando pouco conforme a situação do indivíduo.

As maiores diferenças aparecem nas probabilidades de não responder dado que não respondeu no trimestre inicial: 81,5%, 47,3% e 69,6% para ocupados, desocupados e fora da força no trimestre inicial, respectivamente, mostrando que a não resposta é diferencial de acordo com a situação ocupacional no trimestre inicial. Esse é um achado inédito, o qual indica que análises das estimativas de fluxos e estoques da situação ocupacional podem sofrer viés caso não sejam adotadas estratégias para compensar os efeitos dessa não resposta diferencial. Resultados similares foram obtidos para outros trimestres analisados por Jacob (2021).

#### 4 CONSIDERAÇÕES FINAIS

Embora os resultados apresentados envolvam fluxos brutos para situação ocupacional, a abordagem metodológica e o pacote *surf* podem ser facilmente aplicados a outras distribuições de interesse, tais como situações envolvendo classificações do trabalho principal como formal ou informal, frequência escolar, classes de rendimento etc. Como mencionado, com amostras suficientemente grandes, os procedimentos propostos produzem estimativas consistentes para os parâmetros populacionais, independentemente da bondade do ajuste dos modelos empregados para deduzir estimadores (Särndal, Swensson e Wretman, 1992). Por esta razão, a abordagem implementada no pacote *surf* é particularmente adequada para a produção de estatísticas oficiais sobre fluxos brutos, tais como os da situação ocupacional aqui considerados. O pacote está planejado para ser distribuído gratuitamente através da Comprehensive R Archive Network (CRAN) ainda no primeiro trimestre de 2021.<sup>4</sup>

Algumas limitações do método impõem cuidados do analista ao usar o pacote: zeros na tabulação cruzada do processo observável na amostra podem causar dificuldades para a convergência do algoritmo de ajuste do modelo; e algumas configurações de contagens observáveis podem causar não convergência das estimativas sob os modelos C e D para a não resposta. No entanto, tais situações não ocorreram em nenhuma das análises feitas com o conjunto do total das amostras da PNAD Contínua para trimestres no período entre o 1º trimestre de 2013 e o 1º trimestre de 2019.

No momento, os autores trabalham em dois desenvolvimentos possíveis. O primeiro é um método que combina linearização e reamostragem para estimar variâncias das estimativas dos parâmetros do modelo (Chambers *et al.*, 2012, p. 62), que pode ser necessário nas ocasiões em que o analista não dispuser de todas as informações sobre a estrutura do plano amostral, tendo acesso apenas a pesos de replicação.

4. Enquanto não ocorre essa distribuição, o pacote pode ser encontrado em: <<https://github.com/guilhermejacob/surf>>.

O segundo é a extensão da abordagem mediante formulação de modelos mais adequados para lidar com pesquisas que usam painéis rotativos, permitindo uma formulação estimável mais próxima do modelo geral de Stasny (1987). Tais modelos, ao menos em teoria, permitiriam enfrentar o problema do “viés de grupo de rotação” (Bailar, 1975), que se manifesta como efeitos do tempo de permanência dos indivíduos na pesquisa sobre as respostas, e que pode afetar pesquisas amostrais de painéis rotativos, como é o caso da PNAD Contínua.

## REFERÊNCIAS

- BAILAR, B. A. The effects of rotation group bias on estimates from panel surveys. **Journal of the American Statistical Association**, v. 70, n. 349, p. 23-30, 1975.
- BINDER, D. A. On the variances of asymptotically normal estimators from complex surveys. **International Statistical Review**, v. 51, n. 3, p. 279-292, 1983.
- CHAMBERS, R. L. *et al.* **Maximum likelihood estimation for sample surveys**. Boca Raton: CRC Press, 2012.
- GUTIÉRREZ, H. A. **Modelos para estimar cambios brutos en encuestas rotativas con ausencia de respuesta en diseños de muestreo complejos**. 2014. Tese (Doctorado) – Facultad de Ciencias, Departamento de Estadística, Universidad Nacional de Colombia, Bogotá, Colombia, 2014.
- IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional por Amostra de Domicílios Contínua**: notas técnicas – versão 1.8. Rio de Janeiro, 2020.
- JACOB, G. A. P. **Estimación de fluxos ocupacionais brutos usando a PNAD Contínua**. Dissertação (Mestrado em População, Território e Estatísticas Públicas) – Escola Nacional de Ciências Estatísticas, Fundação Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, 2021.
- LEPKOWSKI, J.; KALTON, G.; KASPRZYK, D. **Weighting adjustments for partial nonresponse in the 1984 SIPP panel** – proceedings of the survey research methods section. Alexandria: American Statistical Association, 1989.
- LUMLEY, T. Analysis of complex survey samples. **Journal of Statistical Software**, v. 9, n. 1, p. 1-19, 2004.
- MOLINA, E. A.; SKINNER, C. J. Pseudo-likelihood and quasi-likelihood estimation for complex sampling schemes. **Computational Statistics & Data Analysis**, v. 13, n. 4, p. 395-405, 1992.
- SÄRNDAL, C.-E.; SWENSSON, B.; WRETMAN, J. **Model assisted survey sampling**. Nova York: Springer-Verlag, 1992.
- STASNY, E. A. Some Markov-chain models for nonresponse in estimating gross labor force flows. **Journal of Official Statistics**, v. 3, n. 4, p. 359-373, 1987.
- TEIXEIRA JÚNIOR, A. E. *et al.* Pesos longitudinais para a Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua). **Mercado de Trabalho: Conjuntura e Análise**, n. 67, out. 2019.



## **Ipea – Instituto de Pesquisa Econômica Aplicada**

### **EDITORIAL**

#### **Chefe do Editorial**

Reginaldo da Silva Domingos

#### **Assistentes da Chefia**

Rafael Augusto Ferreira Cardoso

Samuel Elias de Souza

#### **Supervisão**

Camilla de Miranda Mariath Gomes

Everson da Silva Moura

#### **Editoração**

Aeromilson Trajano de Mesquita

Anderson Silva Reis

Cristiano Ferreira de Araújo

Danilo Leite de Macedo Tavares

Jeovah Herculano Szervinsk Junior

Leonardo Hideki Higa

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

#### **Livraria Ipea**

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo

70076-900 – Brasília – DF

Tel.: (61) 2026-5336

Correio eletrônico: [livraria@ipea.gov.br](mailto:livraria@ipea.gov.br)



## **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

