

## DATAZOOM PAINEL DA PNAD CONTÍNUA

**Aluna: Maria Mittelbach**  
**Orientador: Gustavo Gonzaga**

### **Introdução**

O trabalho consistiu em criar um painel longitudinal da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua). A PNAD Contínua é realizada através de uma amostra de domicílios com um esquema de rotação. Assim, o mesmo domicílio é entrevistado cinco vezes e, por isso, o mesmo indivíduo pode ser entrevistado mais de uma vez. No entanto, a identificação desse indivíduo não é disponibilizada. Utilizando a metodologia de Ribas e Soares (2008), o trabalho a seguir se propõe a criar um código que ajude os usuários do DataZoom na identificação dos indivíduos ao longo do tempo.

Duas formas de identificação dos indivíduos foram criadas. A primeira forma de identificação consiste na forma simples. Nela utilizam-se as informações que deveriam ser fixas ao longo dos trimestres para fazer o match dos indivíduos, como, por exemplo, data de aniversário e ano de nascimento.

No entanto, em Ribas e Soares (2008) se afirma que muitas vezes as pessoas dão as informações erradas ou imprecisas e, por isso, a taxa de identificação em pesquisas amostrais se torna mais baixa. Seguindo a metodologia do artigo, o trabalho a seguir cria uma forma de identificação avançada que permite uma flexibilização dessas informações objetivando aumentar a taxa de identificação dos indivíduos.

### **Pesquisa Nacional por Amostra de Domicílios Contínua**

A Pesquisa Nacional Contínua por Amostra de Domicílios — PNAD Contínua do Instituto Brasileiro de Geografia e Estatística (IBGE), é uma pesquisa domiciliar produzida periodicamente a cada trimestre. O objetivo principal da pesquisa é produzir informações contínuas sobre a inserção e condição da população brasileira na força de trabalho e outros indicadores socioeconômicos. Ela é parte do Sistema Integrado de Pesquisas Domiciliares (SIPD), que se constitui em um modelo de produção de pesquisas amostrais domiciliares. A pesquisa divulga informações como a população em idade de trabalhar, rendimento médio, população ocupada, entre outros indicadores importantes para pesquisas socioeconômicas. Além de ter uma periodicidade menor, a PNAD Contínua tem uma maior abrangência territorial do que a PNAD, pois abrange todo o país com exceção de áreas com características especiais como áreas indígenas, bases militares, orfanatos e outros. Assim, a PNAD Contínua permanece como um instrumento importante em pesquisas socioeconômicas.

#### **a. Principais Indicadores:**

Os principais indicadores de mercado de trabalho que a PNAD Contínua divulga são:

- Pessoas em idade de trabalhar;
- Condição de ocupação na semana de referência;
- Força de trabalho na semana de referência;
- Categoria do emprego do trabalho principal;

- Posição na ocupação do trabalho principal;
- Taxa de atividade;
- Taxa de desocupação;
- Nível de ocupação;
- Nível de desocupação.

**b. Abrangência Geográfica:**

A pesquisa tem como objetivo através da pesquisa domiciliar conseguir representatividade dos seguintes níveis geográficos:

- Brasil;
- Grandes Regiões;
- Unidades da Federação;
- Regiões Metropolitanas, que incluem os municípios das capitais.

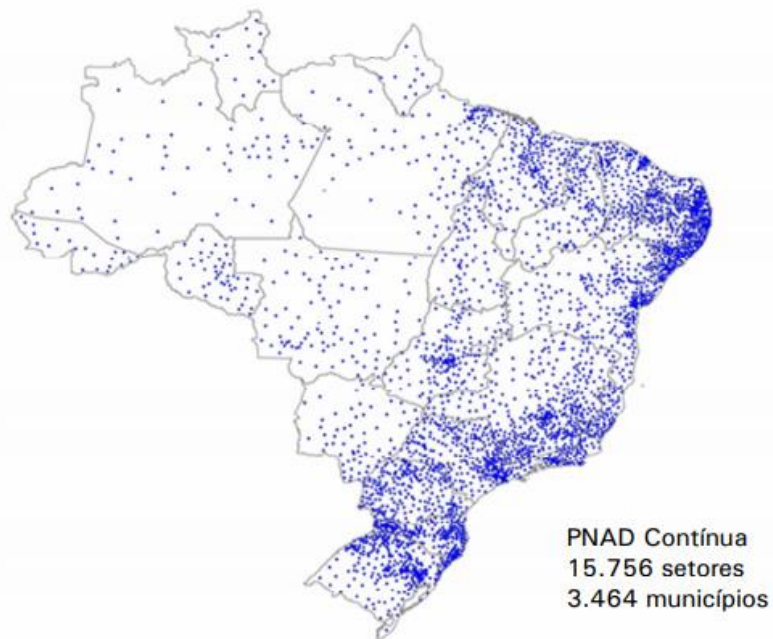


Figura 1. Municípios que contém setores das amostras da PNAD Contínua. Fonte: IBGE (2013).

**c. Rotação da amostra:**

Cada domicílio selecionado é visitado cinco vezes ao longo de cinco trimestres. No entanto, as entrevistas são feitas por um esquema de rotação chamado de 1-2(5). Neste esquema de rotação, o domicílio é entrevistado por um mês e deixa a amostra pelos 2 meses seguintes, e esta sequência é repetida cinco vezes. Assim, eles entram e saem da amostra de forma predeterminada.

Mês	Painel																																					
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	C1	C2	C3	C4	C5	C6		
jan/09	5			4			3			2			1																									
fev/09		5			4			3			2			1																								
mar/09			5			4			3			2			1																							
abr/09				5			4			3			2		1																							
mai/09					5			4			3			2		1																						
jun/09						5			4			3			2		1																					
jul/09							5			4			3			2		1																				
ago/09								5			4			3			2		1																			
set/09									5			4			3			2		1																		
out/09										5			4			3			2		1																	
nov/09											5			4			3			2		1																
dez/09												5			4			3			2		1															
jan/10													5			4			3			2		1														
fev/10														5			4			3			2		1													
mar/10															5			4			3			2		1												
abr/10																5			4			3			2		1											
mai/10																	5			4			3			2		1										
jun/10																		5			4			3			2		1									
jul/10																			5			4			3			2		1								
ago/10																				5			4			3			2		1							
set/10																					5			4			3			2		1						
out/10																						5			4			3			2		1					
nov/10																							5			4			3			2		1				
dez/10																								5			4			3			2		1			

Figura 2. Esquema de rotação da amostra de domicílios: 1-2(5). Fonte: IBGE (2014).

## Metodologia

A metodologia de Ribas e Soares (2008) criada para identificação dos indivíduos na PME foi a base deste trabalho. Assim como a PME, a PNAD Contínua acompanha o mesmo domicílio durante um período de tempo. No caso da PNAD Contínua, cada domicílio selecionado é visitado por cinco vezes, durante cinco trimestres consecutivos. No entanto, assim como no caso da PME, existem algumas dificuldades na hora de construir um painel.

Os dois pontos de atritos mais comumente reportados são a mobilidade geográfica das pessoas e a recusa de entrevistas. Porém, Ribas e Soares (2008) afirmam que existe um outro tipo de atrito ao se tentar construir um painel que seria a imprecisão nas respostas da entrevista. Para tentar reduzir esse atrito, os autores propõem um tipo avançado de pareamento que leve em conta possíveis erros nas respostas.

No emparelhamento básico, levamos em consideração as características individuais que são constantes no tempo para determinar se duas pessoas entrevistadas em trimestres diferentes na mesma casa são a mesma pessoa. Essas características são: o dia, mês e ano de nascimento, e sexo (Lopes, 2002). No entanto, Ribas e Soares (2008) afirmam que, no caso de pesquisas amostrais domiciliares, essas características são muito sensíveis e, assim, ao utilizar essa forma de emparelhamento se obtém taxas de identificação baixas. Por isso, os autores criaram uma forma de emparelhamento que considera a possibilidade de pequenas mudanças nas características através do tempo. Inicialmente, os domicílios passam pelo emparelhamento básico, no entanto os indivíduos que não foram identificados dessa forma irão passar pelo emparelhamento avançado. Nesse segundo critério, se avalia não apenas a exatidão das respostas, mas a proximidade delas.

## Resultados

Observando a tabela 1, nota-se um aumento na identificação de indivíduos ao realizar o emparelhamento avançado. O emparelhamento aumenta em todos os anos para os indivíduos e se mantém constante para os domicílios, uma vez que esses têm um bom emparelhamento com dados da PNAD Contínua. O emparelhamento de domicílios não é perfeito, por conta da saída de domicílios da amostra, recusa de entrevista e outros atritos.

Pelas Tabelas 2 e 3, nota-se que os atritos diminuem em relação ao painel avançado, corroborando a ideia de que ele ajuda na identificação dos indivíduos.

Basic matching criteria														
time interval	Painel 1		Painel 2		Painel 3		Painel 4		Painel 5		Painel 6		Painel 7	
	House.	Indiv.	House.	Indiv.	House.	Indiv.	House.	Indiv.	House.	Indiv.	House.	Indiv.	House.	Indiv.
1	67.7	55.1	91.0	67.4	92.6	66.0	93.4	64.7	93.8	56.1	93.7	57.2	70.5	48.7
2	41.3	30.7	84.0	54.4	87.1	54.7	88.2	54.4	89.1	48.0	88.6	49.2	44.7	28.9
3	18.8	12.9	77.7	45.1	82.1	46.4	83.3	46.6	84.6	42.1	84.0	43.3	21.0	12.7
4			69.4	35.8	75.2	37.7	77.1	38.7	78.5	36.2	78.2	37.7		
Advanced matching criteria														
time interval	Painel 1		Painel 2		Painel 3		Painel 4		Painel 5		Painel 6		Painel 7	
	House.	Indiv.	House.	Indiv.	House.	Indiv.	House.	Indiv.	House.	Indiv.	House.	Indiv.	House.	Indiv.
1	67.7	61.0	91.0	78.1	92.6	79.1	93.4	81.3	93.8	83.4	93.7	84.2	70.5	65.2
2	41.3	35.0	84.0	65.3	87.1	67.6	88.2	70.0	89.1	73.1	88.6	73.9	44.7	39.3
3	18.8	15.2	77.7	55.8	82.1	59.2	83.3	61.5	84.6	65.1	84.0	66.0	21.0	17.6
4			69.4	46.6	75.2	50.5	77.1	53.4	78.5	56.9	78.2	58.3		

Tabela 1. Sobreposição da amostra na PNAD Contínua entre pares de meses.

	time interval	Individuals		
		Households	In matched households	
			Total	
Painel 1	1	31.6	44.9	20.2
	2	39.6	44.3	11.2
	3	54.4	57.9	13.8
	4			
Painel 2	1	5.1	32.6	24.2
	2	6.6	19.3	12.1
	3	12.3	17.1	10.9
	4	10.7	20.5	13.4
Painel 3	1	3.9	34.0	27.7
	2	5.0	17.2	11.7
	3	10.0	15.2	10.5
	4	8.4	18.7	13.3
Painel 4	1	3.6	35.3	30.0
	2	4.7	16.0	10.8
	3	9.3	14.3	9.8
	4	7.5	16.9	12.0
Painel 5	1	3.4	43.9	40.6
	2	4.3	14.3	9.4
	3	8.5	12.4	8.2
	4	7.2	13.8	8.8
Painel 6	1	3.5	42.8	39.3
	2	4.6	14.0	9.0
	3	8.7	12.0	7.6
	4	6.9	12.9	8.0
Painel 7	1	29.1	51.3	39.3
	2	37.0	40.6	9.5
	3	53.1	56.1	9.9
	4			

Tabela 2. Perdas entre pares de trimestres, seguindo os critérios básicos de emparelhamento (em %)

		Percentage of individuals		
		Friction (not recovered)		
		time interval	In matched households	Recovered
Painel 1	1	39.0	11.0	5.8
	2	42.6	7.9	1.7
	3	56.5	8.4	1.4
	4			0.0
Painel 2	1	21.9	12.5	10.7
	2	16.4	8.9	2.9
	3	14.5	8.0	2.7
	4	16.6	8.6	3.9
Painel 3	1	20.9	13.1	13.1
	2	14.5	8.6	2.7
	3	12.5	7.5	2.7
	4	14.6	8.6	4.0
Painel 4	1	18.7	11.3	16.5
	2	13.8	8.3	2.2
	3	12.1	7.3	2.2
	4	13.2	7.8	3.7
Painel 5	1	16.6	9.5	27.4
	2	12.4	7.1	1.9
	3	10.9	6.5	1.5
	4	12.6	7.3	1.3
Painel 6	1	15.8	8.6	26.9
	2	12.2	6.7	1.9
	3	10.8	6.1	1.3
	4	11.7	6.5	1.2
Painel 7	1	34.8	8.9	16.5
	2	39.8	7.2	0.8
	3	55.2	7.2	0.9
	4			0.0

Tabela 3. Perdas entre pares de trimestres, seguindo os critérios avançados de emparelhamento (em %).

## Conclusões

Este trabalho teve como objetivo criar um código para ajudar na identificação de indivíduos na PNAD Contínua, utilizando a metodologia de Ribas e Soares (2008). Para isso, criou-se um painel com um emparelhamento básico que utilizava data de nascimento e sexo como as variáveis de pareamento fixas ao longo do tempo. No entanto, a realidade de pesquisas por amostra é que muitas dessas variáveis podem vir a mudar por erros ou outros fatores e, assim, o emparelhamento avançado é feito a fim de aumentar a taxa de identificação ao permitir maior flexibilidade nas respostas. Com esse novo tipo de emparelhamento, há um aumento na taxa de identificação em todos os anos da PNAD Contínua e uma diminuição na taxa de atrito.

## **Referências**

- 1 - RIBAS, Rafael Perez; SOARES, Sergei Suarez Dillon. Sobre o painel da Pesquisa Mensal de Emprego (PME) do IBGE. Texto para discussão, 2008.
- 2 - IBGE. Notas Técnicas. Instituto Brasileiro de Geografia e Estatística, 2013
- 3 - Integrado de Pesquisas Domiciliares: amostra mestra 2010 e amostra da PNAD contínua. Rio de Janeiro: IBGE, 2014.