

É possível gerar estimativas conjunturais a partir de dados longitudinais extraídos da Pnad Contínua?*

Gustavo Plínio P. Monteiro**

Resumo: o texto apresenta uma metodologia de uso da Pnad Contínua como fonte de informações. São abordados soluções para três problemas que surgem nessa tarefa: (1) como identificar os mesmos indivíduos entre um trimestre e outro; (2) como ajustar os pesos amostrais das pessoas na subamostra de dados longitudinais da Pnad Contínua; (3) como estimar os erros amostrais dessa subamostra. O texto apresenta uma discussão sobre como superar esses três problemas, visando à obtenção de uma subamostra com dados longitudinais que sejam representativos e, ao mesmo tempo, que possam ser facilmente obtidos por pesquisadores das ciências sociais, possibilitando a realização de estudos conjunturais regulares, que acompanhem *pari passu* as divulgações trimestrais da Pnad Contínua.

Palavras-chave: Pesquisa nacional por amostra de domicílios; Estudo longitudinal; Mercado de trabalho.

O termo “dados longitudinais” refere-se a um tipo especial de disposição de dados, em que as mesmas unidades em corte transversal (indivíduos, empresas ou domicílios, por exemplo) são acompanhadas ao longo de um período de tempo. Uma das principais vantagens desse tipo de banco de dados é que ele é mais adequado para o estudo de padrões de mudanças e transições, permitindo o acompanhamento da evolução de certos fenômenos.

Dados longitudinais sobre as pessoas investigadas pela Pesquisa Nacional por Amostra de Domicílios (Pnad) Contínua podem fornecer importantes informações sobre a dinâmica do mercado de trabalho. Essa base pode ser utilizada

* O autor agradece a ajuda dos técnicos do IBGE Marina Águas e Maurício Lila Freitas; da Prof. Denise Britz Silva e de Antônio Teixeira Júnior, ambos da ENCE/IBGE; da economista do DEPEC/Bradesco, Ana Maria Bonomi Barufi; e dos técnicos do DIEESE Camila Ikuta e Edgard Fusaro. Eventuais erros, inconsistências ou omissões são de responsabilidade do autor.

** Técnico do DIEESE. E-mail: gustavopmonteiro@gmail.com

para estimar a duração do período de desemprego, calcular a rotatividade no mercado de trabalho, acompanhar a mobilidade ocupacional dos trabalhadores ou analisar a transição entre diferentes condições de ocupação das pessoas (ingresso e saída do mercado de trabalho, por exemplo)¹.

O objetivo deste texto é apresentar uma metodologia que auxilie pesquisadores das ciências sociais, interessados na utilização da Pnad Contínua como fonte de informações longitudinais representativas, em seus estudos de conjuntura socioeconômica.

A Pnad Contínua

A Pnad Contínua fornece uma série de informações socioeconômicas do país, possibilitando o acompanhamento de diversos indicadores conjunturais de trabalho e rendimento. A pesquisa é realizada através de uma amostra probabilística de domicílios. São investigados, trimestralmente, cerca de 211 mil domicílios em aproximadamente 16 mil setores censitários.

Diferentemente da Pnad, a Pnad Contínua tem caráter longitudinal, uma vez que o mesmo *domicílio* é entrevistado cinco vezes, com um intervalo de três meses entre uma entrevista e outra. Portanto, entre a primeira e a última entrevista de cada domicílio, transcorre o período de um ano completo.

Para possibilitar que os dados da amostra da Pnad Contínua sejam transformados em um banco de dados longitudinais (ou, uma subamostra de dados longitudinais) é necessário enfrentar ao menos três problemas, que serão abordados na próxima seção: (1) como identificar os mesmos indivíduos entre um trimestre e outro; (2) como fazer o ajuste dos pesos amostrais das pessoas na subamostra de dados longitudinais da Pnad Contínua; (3) como estimar os erros amostrais dessa subamostra.

O texto apresenta uma discussão sobre como superar esses três problemas, visando à obtenção de uma subamostra com dados longitudinais que sejam

¹ Alguns estudos recentes já utilizaram os dados longitudinais da Pnad Contínua, como AMORIM e CORSEIUL (2016); MOREIRA, CORSEIUL e FOGUEL (2016); REIS (2017); e BCB (2016).

representativos e que possam ser facilmente obtidos, o que possibilita a realização de estudos conjunturais regulares, que acompanhem *pari passu* as divulgações trimestrais da Pnad Contínua.

Características da subamostra de dados longitudinais da Pnad Contínua

Esta seção apresentará três problemas que surgem em estudos que buscam utilizar a Pnad Contínua como fonte de informações longitudinais, bem como um método de enfrentamento desses problemas. Para entender melhor os impactos desses problemas, esta seção também aborda alguns resultados obtidos a partir de uma subamostra de dados longitudinais, que foi extraída do primeiro e do segundo trimestres de 2017 da Pnad Contínua, aplicando a metodologia aqui apresentada.

Os três problemas que surgem são, em resumo:

1. Como identificar as mesmas pessoas investigadas em diferentes pontos no tempo?
2. Como ajustar os pesos amostrais da subamostra obtida, após as perdas que a amostra original inevitavelmente sofre?
3. Como avaliar se o tamanho da subamostra resultante é adequado para que possam ser geradas estimativas para o total da população, a partir dessa base de dados? Ou seja, qual a dimensão dos erros amostrais (ou qual a precisão das estimativas obtidas) dessa subamostra com dados longitudinais?

Chave identificadora dos indivíduos

A Pnad Contínua não permite que o acompanhamento das mesmas pessoas, ou seja, dos mesmos *indivíduos* seja feito de forma direta. Em primeiro lugar, porque a pesquisa investiga os mesmos *domicílios* de forma recorrente, mas nada garante que os *indivíduos* investigados pela primeira entrevista em determinado domicílio sejam encontrados residindo no mesmo domicílio nas próximas quatro entrevistas (perda por não resposta). Alguns indivíduos podem sair da amostra entre um trimestre e outro, seja porque se mudaram, seja porque não responderam ao entrevistador em uma das visitas subsequentes que este fez ao

domicílio. Em segundo lugar, não há uma maneira direta de acompanhar as mesmas pessoas entre um trimestre e outro, porque o IBGE fornece apenas a chave de identificação dos domicílios, mas não divulga a chave de identificação dos moradores dentro de cada domicílio.

Para resolver o primeiro problema, de perda por não resposta, optou-se por utilizar somente as informações de trimestres consecutivos, ao invés de acompanhar um período mais longo – um ano por exemplo. Espera-se que isso reduza as perdas de informação por não resposta. Se essas perdas forem não aleatórias, isto é, ocorrerem devido a um padrão de comportamento dos indivíduos, isso acarretaria em um viés de seleção dentro da amostra. Seguindo o procedimento utilizado por Moreira, Amorim e Corseiul (2016), adotou-se o pressuposto de que os problemas decorrentes de perdas não aleatórias por não resposta serão mitigados se se considerar um período mais curto entre as entrevistas: um trimestre. Segundo Pizzot *et al.* (2017), essa subamostra longitudinal da Pnad Contínua, que abrange os dados obtidos em entrevistas realizadas em dois trimestres consecutivos, pode ser considerada representativa. Assim, vai-se acompanhar as respostas dos indivíduos entre o primeiro (T1) e o segundo trimestre (T2) de cada ano.

Para resolver o segundo problema, isto é, o da falta de uma chave de identificação individual para os moradores dos domicílios investigados, utilizar-se-á a estratégia de combinar as informações de moradia, sexo e data de nascimento dos moradores. Portanto, as variáveis utilizadas na identificação dos domicílios e dos indivíduos na amostra, para gerar a subamostra de dados longitudinais são as seguintes²:

- Variáveis que identificam os domicílios:
 - *UPA*: Unidade Primária de Amostragem
 - *V1008*: Número de seleção do domicílio
 - *V1014*: Painel (grupo da amostra)

²Baseado no trabalho de MOREIRA, AMORIM e CORSEIUL (2016). As variáveis de identificação dos domicílios são apresentadas no documento divulgado pelo IBGE através do endereço eletrônico: <ftp://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Trimestral/Microdados/Documentacao/Leia_me.pdf>.

- Variáveis que identificam os indivíduos:
 - V2007: Sexo
 - V2008: Dia de nascimento
 - V20081: Mês de nascimento
 - V20082: Ano de nascimento

Sem nenhuma perda, a subamostra deveria conter aproximadamente 80% do tamanho da amostra original, uma vez que os domicílios que realizam a sua quinta entrevista no primeiro trimestre (T1) estão em sua última entrevista e, portanto, não aparecerão na pesquisa realizada no trimestre seguinte (T2). De forma análoga, os domicílios que realizaram sua primeira entrevista em T2 não terão aparecido na pesquisa em T1, de forma que um quinto (20%) da amostra de cada um dos dois trimestres será descartada para a obtenção de um banco de dados longitudinais.

QUADRO 1

Tamanho da amostra antes e após perdas (subamostra de dados longitudinais da Pnad Contínua) – Brasil - 1º e 2º trimestres de 2012 a 2017 (em n^{os} absolutos)

Tamanho da amostra/subamostra	2012	2013	2014	2015	2016	2017
Tamanho total da amostra (T1 + T2)	1.133.889	1.137.489	1.146.102	1.147.016	1.137.585	1.140.692
Tamanho da amostra no 1º trimestre (T1)	566.873	564.773	573.480	572.982	566.932	572.379
Tamanho da amostra no 2º trimestre (T2)	567.016	572.716	572.622	574.034	570.653	568.313
Eliminação da 5ª entrevista em T1 e da 1ª entrevista em T2 (E)	227.502	227.931	228.483	227.014	225.710	229.173
% eliminados (~20%) (E/(T1+T2))	20,1	20,0	19,9	19,8	19,8	20,1
Subamostra após eliminação ((T1+T2) - E)	906.387	909.558	917.619	920.002	911.875	911.519
Perda por idade ignorada (PII)	37.701	35.275	42.640	44.045	83.147	106.784
Perda por não resposta (PPNR)	175.354	142.107	181.379	173.351	101.078	90.483
Subamostra após perdas (T1+T2 - E - PII - PPNR)	693.332	732.176	693.600	702.606	727.650	714.252
% da subamostra/amostra ((T1+T2 - E - PII - PPNR)/(T1+T2))	61,1	64,4	60,5	61,3	64,0	62,6

Fonte: IBGE. PNAD Contínua. Elaboração própria.

Além dessas perdas, cerca de um quarto da subamostra remanescente se perde, seja porque não foi possível identificar a data de nascimento do indivíduo (perda por idade ignorada), seja porque o mesmo indivíduo não foi encontrado no

domicílio, através dos critérios adotados, nos dois trimestres investigados (perda por não resposta).

Assim, após a identificação dos indivíduos investigados, a subamostra resultante para os anos entre 2012 e 2017 é cerca de 40% menor do que a amostra total da pesquisa. O Quadro 1 contém uma síntese do tamanho da amostra inicial e final, após todas as perdas acima citadas, para os bancos longitudinais obtidos a partir do primeiro e do segundo trimestre de cada ano, entre 2012 e 2017.

A próxima seção discute como ajustar os pesos da subamostra de modo a compensar essas perdas.

Ajuste dos pesos amostrais da subamostra

A Pnad Contínua atribui para cada morador um fator de expansão, chamado também de peso amostral, que possibilita que as estimativas obtidas através da amostra probabilística da pesquisa sejam expandidas de forma a apresentar estimativas que representem toda a população abrangida pela pesquisa.

Para definir os pesos das observações de suas amostras, o IBGE calibra as estimativas obtidas na Pnad Contínua utilizando a projeção da população das unidades da Federação por Sexo e Idade, calculada pelo próprio IBGE³. Essa projeção considera o crescimento populacional ocorrido entre os Censos Demográficos, levando também em conta a formulação de hipóteses sobre o comportamento futuro de parâmetros demográficos, referentes às taxas de fecundidade, de mortalidade e a migração. Assim, as estimativas de tamanho da população obtidas pela Pnad Contínua são ajustadas, de modo que correspondam às projeções populacionais calculadas pelo IBGE para cada uma das 77 regiões geográficas correspondentes aos domínios de estimação da pesquisa⁴.

Existem métodos sofisticados, que calculam os pesos de amostras longitudinais, utilizando modelagem de propensão de resposta por regressão logística e, dessa maneira, calculam um peso longitudinal que corrige tanto as

³ IBGE, 2013.

⁴ IBGE, 2017.

perdas quanto o viés da pesquisa⁵. Neste texto, porém, será apresentado um método mais simples, mas que, apesar de menos refinado, pode ser replicado de forma mais fácil e regular, permitindo o acompanhamento conjuntural da economia e do mercado de trabalho⁶.

Para tanto, devem ser calculados novos pesos para as pessoas investigadas que permaneçam na subamostra de dados longitudinais, de forma a compensar as perdas ocorridas em relação à amostra original, conforme apresentadas na seção anterior. Os pesos de expansão dessa subamostra consideram os respondentes no primeiro trimestre e devem ser ajustados de acordo com as perdas, devido a problemas de identificação entre um trimestre e outro e devido a não resposta no segundo trimestre.

O cálculo dos novos pesos é feito em duas etapas:

1. A primeira etapa ajusta os pesos de forma a compensar as perdas em relação à amostra original, apresentadas no Quadro 1;
2. Na segunda etapa, os pesos da subamostra devem ser ajustados para que as estimativas da população da subamostra, por sexo e faixa etária, correspondam às estimativas, por sexo e faixa etária, obtidas para a amostra original da Pnad Contínua, em cada um dos 77 domínios de estimação (variável *posest*).

Conforme exposto acima, na segunda etapa, o ajuste é feito utilizando não apenas as informações geográficas, como também de sexo e faixa etária da população. Esse passo é importante porque a subamostra longitudinal poderia potencializar um viés que já é inerente à Pnad Contínua. Trata-se do viés de disponibilidade nas entrevistas. Dito de maneira direta: “(...) *mulheres e idosos estão*

⁵ TEIXEIRA JUNIOR et al., 2019

⁶ Neste texto, será apresentada uma metodologia de ajuste dos pesos semelhante à que o IBGE empregou para ajustar os pesos do Suplemento de Mobilidade Sócio Ocupacional da Pnad de 2014 (IBGE, 2016).

*sobre representados na amostra, pois são encontrados com maior facilidade em seus domicílios, em comparação a homens jovens*⁷.

A elaboração da subamostra de dados longitudinais exige a combinação de duas entrevistas consecutivas, realizadas com um intervalo de três meses entre elas. Se a probabilidade de ser entrevistado uma vez já é menor para um homem jovem, espera-se que essa probabilidade fique ainda menor, quando se considera a chance de um homem jovem ser entrevistado (ou captado nas entrevistas) por dois trimestres consecutivos.

Vasconcelos e Silva (2005) recomendam a calibração por sexo e idade, visto que essa é *“(...) a melhor alternativa técnica para lidar com o mais típico viés de coleta e seleção*⁸.

Dessa forma, o ajuste dos pesos dos indivíduos da subamostra de dados longitudinais, que foram investigados no primeiro e no segundo trimestre, foi calculado da seguinte forma:

$$w_j^* = w_j \times \frac{n_d}{n_d^*} \times \frac{P_{dsi}}{\hat{P}_{dsi}}$$

Em que:

w_j^* = novo peso em T1, calculado para cada indivíduo j da amostra;

w_j = peso para cada indivíduo j da amostra da Pnad Contínua (V1028), em T1;

n_d = número total de pessoas entrevistadas na área geográfica representada pelo domínio de estimação d , em T1;

n_d^* = número total de pessoas na subamostra, após perdas, na área geográfica representada pelo domínio de estimação d , em T1;

P_{dsi} = estimativa populacional produzida pelo IBGE para o domínio de estimação d , o sexo s e a faixa etária⁹ i , na data de referência¹⁰;

⁷ RUIZ e SILVA, 2014, p. 2. Convém, ainda, destacar que, além dos homens jovens, a população na força de trabalho e as pessoas que moram sozinhas também são sub-representados na pesquisa (PIZZOTT *et al.*, 2017).

⁸ *Apud* TEIXEIRA JUNIOR, 2016, p. 8

⁹ As faixas etárias utilizadas foram: menos de 14 anos; entre 14 e 17 anos; entre 18 e 24 anos; entre 25 e 29 anos; entre 30 e 44 anos; entre 45 e 64 anos e, 65 anos ou mais.

\hat{P}_{dsi} = estimativa populacional obtida com os dados da subamostra para o domínio de estimação d , o sexo s e a faixa etária i , em T1.

A próxima seção considera como estimar a precisão da subamostra, ou seja, como avaliar se o tamanho da subamostra é suficiente para que se possa usá-la para gerar estimativas para toda a população.

Precisão das estimativas

Para avaliar se o número de observações que permaneceram na subamostra de dados longitudinais, após todas as perdas apresentadas na seção 3.1, tem um tamanho adequado para que possam ser geradas, a partir dessa subamostra, estimativas para toda a população, é preciso que os erros amostrais dessa base sejam dimensionados. Em outras palavras, é necessário que se calcule qual é a precisão das estimativas obtidas a partir da subamostra.

A subamostra de dados longitudinais da Pnad Contínua está sujeita a dois tipos de erros: (1) os erros de amostragem, oriundos de fatores aleatórios e (2) os erros alheios à amostragem, que têm origem não probabilística.

Os erros do primeiro tipo, por serem aleatórios, não levam à necessidade de nenhum tipo de ajuste. Já entre os erros do segundo tipo, estão os erros decorrentes de perdas não aleatórias por não resposta ou por não identificação e, por não serem aleatórios, podem resultar em viés na subamostra. Embora seja praticamente impossível isolar os diferentes tipos de erros, existem procedimentos para ajustar os resultados levando em conta uma estimativa dos erros não aleatórios. Esses métodos, porém, exigem análises mais complexas e processamentos computacionais sofisticados.

Conforme destacado na seção 3.1, adotou-se o pressuposto de que restringir a subamostra aos dois primeiros trimestres de um mesmo ano é uma forma de prevenir possíveis problemas que um padrão não aleatório de não respostas

¹⁰ Dia 15 do segundo mês do trimestre. No caso do primeiro trimestre, a data de referência é o dia 15 de fevereiro.

pode ocasionar, já que, num período de três meses, as perdas por não resposta serão menores e a amostra de respondentes pode ser considerada representativa.

Por isso, neste texto, os resultados referem-se apenas às estimativas de erros do primeiro tipo, isto é, dos erros amostrais aleatórios. Esses erros são expressos pelos coeficientes de variação de estimativas geradas a partir da subamostra de dados longitudinais.

O coeficiente de variação (CV) é uma medida de dispersão que mede a homogeneidade dos dados em relação à média. Ele é definido como a razão percentual entre o desvio padrão S e a média amostral:

$$CV = \frac{S}{\bar{x}} \times 100$$

Se o CV estiver muito elevado, isso significa que o conjunto de dados é muito heterogêneo. No caso da Pnad Contínua, uma estimativa que apresente um CV elevado pode indicar que a amostra não tem tamanho suficiente para que se faça essa estimação com um grau de precisão aceitável.

Por isso, foram calculados os coeficientes de variação para algumas estimativas obtidas a partir da subamostra de dados longitudinais, resultante da junção do primeiro e do segundo trimestres de 2017, com o intuito de avaliar o grau de confiabilidade dessa subamostra¹¹. Em outras palavras, buscam-se evidências de que a subamostra tem tamanho suficiente para gerar estimativas com erros de amostragem (medidos pelos coeficientes de variação) não muito elevados.

A Tabela 1, anexa, mostra o tamanho da amostra e os respectivos coeficientes de variação tanto da Pnad original quanto da subamostra com dados longitudinais, ambas do primeiro trimestre de 2017.

¹¹ Para o cálculo das estimativas de variância, utilizou-se o método do Conglomerado Primário. A estimação da variância adotada (por linearização de Taylor) pode ser utilizada tanto para resultados de totais populacionais quanto para valores representados por razões entre totais de duas características, como, por exemplo, a taxa de desocupação e a taxa de participação (FREITAS e ANTONACI, 2014). Para mais detalhes sobre o método do Conglomerado Primário e linearização de Taylor, ver “Análise de Dados Amostrais Complexos”, disponível em: <<https://djalmapessoa.github.io/adac/index.html>>.

Nota-se que o CV de todas as estimativas apresentadas na Tabela 1 está abaixo dos 10%. Se se adotar a convenção de que o dado é considerado heterogêneo para valores de CV acima de 15%, as estimativas da Tabela 1 indicam que o tamanho da subamostra de dados longitudinais, para o primeiro trimestre de 2017, permite gerar estimativas com erros amostrais relativamente baixos para os níveis de desagregação geográfica apresentados¹². Mesmo assim, recomenda-se que todas as estimativas geradas sejam apresentadas juntamente com seus respectivos coeficientes de variação, uma vez que populações pequenas ou concentradas geograficamente podem gerar estimativas com elevados erros amostrais. Esse procedimento é especialmente importante no caso da subamostra com dados longitudinais, visto que seu tamanho reduzido contribui para a limitação da precisão das suas estimativas, o que é indicado pelos coeficientes de variação mais elevados do que os apresentados pela Pnad Contínua original, conforme a Tabela 1.

Os erros amostrais, mesmo que moderados, não podem ser completamente eliminados, o que leva à pergunta se é possível assumir que a subamostra de dados longitudinais possui características populacionais semelhantes às da amostra original da Pnad Contínua. A Tabela 2, anexa, apresenta o total de pessoas da Pnad e da subamostra, bem como as estimativas ajustadas, segundo uma série de características individuais, de trabalho e rendimento, para o primeiro trimestre de 2017.

As duas amostras – a da Pnad Contínua original e a da subamostra de dados longitudinais – fornecem estimativas semelhantes para o conjunto de características apresentadas. Destacam-se os valores da taxa de desocupação e de participação, que ficaram muito próximos nas duas amostras, apresentando diferenças apenas a partir da segunda casa decimal.

¹² O IBGE utiliza o coeficiente de variação de 15% como nível de precisão máximo para a publicação sem ressalvas de suas estimativas. FÁVERO et al. (2009) considera coeficientes acima de 30% como heterogêneos.

Considerações finais

Este texto discutiu os impactos e possíveis abordagens de três questões suscitadas durante a montagem de uma base de dados longitudinais a partir da Pnad Contínua: (1) como identificar os mesmos indivíduos entre um trimestre e outro na Pnad Contínua; (2) como ajustar os pesos amostrais, após as perdas por não resposta ou não identificação; (3) como estimar os erros amostrais da subamostra de dados longitudinais dessa base, com o objetivo de avaliar se o tamanho dessa subamostra é adequado ou não para que sejam geradas estimativas passíveis de serem extrapoladas para toda a população, a partir dessa base de dados. Quanto a seu viés, pelo menos para os dados do primeiro trimestre de 2017, é razoável assumir que essa subamostra de dados longitudinais possui características populacionais semelhantes às da amostra original da Pnad Contínua.

As soluções e ressalvas apontadas para cada uma das questões levantadas, bem como a análise dos resultados para a subamostra do primeiro trimestre de 2017, levam à conclusão de que a adoção da metodologia aqui apresentada torna a Pnad Contínua uma fonte adequada para a construção de um banco de dados longitudinais, com o objetivo de realizar análises conjunturais básicas do mercado de trabalho. Ou seja, é possível gerar estimativas conjunturais a partir de dados longitudinais extraídos da Pnad Contínua. Porém, é recomendável que, para todas as estimativas geradas a partir de uma subamostra de dados longitudinais da Pnad Contínua, sejam apresentados os respectivos erros de amostragem, conforme expressos pelos coeficientes de variação, o que possibilita a avaliação da precisão dessas estimativas.

Anexos

TABELA 1. Pessoas na amostra, estimativas expandidas e coeficiente de variação da Pnad e da subamostra com dados longitudinais - Brasil, Unidades da Federação e Regiões Metropolitanas - 1º trimestre de 2017

Nível geográfico	PNAD			Subamostra de dados longitudinais		
	Pessoas na amostra	Estimativas expandidas	CV (em %)	Pessoas na amostra	Estimativas expandidas	CV (em %)
Rondônia	9.865	1.789.370	5,0	4.679	1.789.370	8,0
Acre	11.860	811.713	3,0	6.457	811.713	3,9
Amazonas	16.836	3.893.806	2,7	10.668	3.893.806	3,2
RM de Manaus (AM)	9.103	2.470.004	3,6	5.761	2.470.004	4,1
Roraima	6.806	461.881	3,9	3.628	461.881	4,6
Pará	22.185	8.292.668	2,1	14.792	8.292.668	2,6
RM de Belém (PA)	6.646	2.239.559	3,8	4.393	2.239.559	4,8
Amapá	5.151	786.059	5,9	3.349	786.059	7,6
RM de Macapá (AP)	4.060	586.416	7,1	2.731	586.416	9,0
Tocantins	8.723	1.531.046	3,3	5.041	1.531.046	4,5
Maranhão	39.339	6.946.861	1,6	22.159	6.946.861	2,3
RM de Grande SãoLui	7.343	1.437.911	4,2	2.751	1.437.911	7,7
Piauí	13.476	3.217.446	3,3	9.903	3.217.446	3,8
RIME da Grande Tere	6.011	1.203.215	7,0	4.195	1.203.215	8,2
Ceará	30.010	8.999.542	1,6	19.857	8.999.542	1,8
RM de Fortaleza (CE)	10.253	3.903.624	2,6	6.655	3.903.624	3,2
Rio Grande do Norte	11.968	3.495.629	2,5	8.037	3.495.629	3,1
RM de Natal (RN)	5.028	1.523.326	4,2	3.207	1.523.326	5,3
Paraíba	15.380	3.992.609	2,2	10.668	3.992.609	2,6
RM de João Pessoa (F	5.557	1.313.975	3,4	3.659	1.313.975	4,1
Pernambuco	22.580	9.391.431	1,9	11.946	9.391.431	2,9
RM de Recife (PE)	8.100	3.956.031	2,1	3.308	3.956.031	5,3
Alagoas	22.769	3.362.943	1,5	9.602	3.362.943	3,1
RM de Maceió (AL)	7.673	1.272.674	2,3	2.466	1.272.674	6,3
Sergipe	11.199	2.279.838	2,1	6.940	2.279.838	2,8
RM de Aracaju (SE)	4.034	946.626	3,3	2.323	946.626	4,7
Bahia	28.059	15.299.109	2,1	17.763	15.299.109	2,2
RM de Salvador (BA)	6.065	4.003.035	3,5	4.000	4.003.035	3,7
Minas Gerais	41.825	21.064.989	1,3	30.463	21.064.989	1,5
RM de Belo Horizont	10.370	5.301.577	2,2	7.409	5.301.577	2,5
Espírito Santo	21.007	3.996.214	1,6	11.859	3.996.214	2,2
RM de Grande Vitória	9.885	1.950.359	2,3	4.986	1.950.359	3,3
Rio de Janeiro	38.858	16.693.068	1,4	21.777	16.693.068	1,7
RM de Rio de Janeiro	26.069	12.247.817	1,8	14.232	12.247.817	2,0
São Paulo	44.850	44.976.627	1,8	29.864	44.976.627	1,9
RM de São Paulo (SP)	19.046	21.342.630	2,5	12.662	21.342.630	2,2
Paraná	29.852	11.282.540	1,7	19.452	11.282.540	2,0
RM de Curitiba (PR)	8.467	3.506.959	2,8	5.197	3.506.959	3,7
Santa Catarina	34.031	6.954.562	1,4	22.172	6.954.562	1,8
RM de Florianópolis (4.522	1.007.028	3,8	2.433	1.007.028	5,8
Rio Grande do Sul	29.583	11.297.180	1,4	20.449	11.297.180	1,8
RM de Porto Alegre (10.075	4.230.930	2,3	6.499	4.230.930	3,3
Mato Grosso do Sul	12.119	2.634.995	2,6	8.244	2.634.995	2,9
Mato Grosso	14.422	3.280.064	2,3	8.696	3.280.064	3,0
RM de Vale do Rio Cu	3.859	891.716	3,2	2.298	891.716	4,3
Goiás	19.374	6.747.931	2,3	12.109	6.747.931	2,6
RM de Goiânia (GO)	7.293	2.480.346	3,8	4.542	2.480.346	4,3
Distrito Federal	10.252	3.012.463	2,5	6.552	3.012.463	3,2
Brasil	572.379	206.492.584	-	357.126	206.492.584	-

Fonte: IBGE. PNAD Contínua. Elaboração própria

Nota: (1) Região Administrativa Integrada de Desenvolvimento da Grande Teresina (PI)

TABELA 2. Pessoas na amostra e estimativas expandidas para a população, segundo características socioeconômicas - Brasil - 1º trimestre de 2017

Características socioeconômicas	PNAD				Subamostra			
	Pessoas na amostra	Estimativas expandidas			Pessoas na amostra	Estimativas expandidas		
		Em 1.000 pessoas	Em %	CV		Em 1.000 pessoas	Em %	CV
Escolaridade								
Sem instrução	75.637	22.862	11,8	0,8	44.994	21.839	11,3	1,0
Fundamental incompleto	193.672	64.895	33,5	0,6	121.985	64.462	33,3	0,7
Fundamental completo	46.896	17.242	8,9	0,9	29.054	17.134	8,9	1,2
Médio incompleto	35.300	12.945	6,7	1,0	21.684	13.044	6,7	1,2
Médio completo	115.304	45.927	23,7	0,7	72.082	46.634	24,1	0,8
Superior incompleto	19.695	7.900	4,1	1,4	12.610	8.258	4,3	1,6
Superior completo	48.931	21.779	11,3	1,7	32.211	22.146	11,4	1,6
Cor/raça								
Não-negro	225.290	91.414	44,3	0,8	146.599	92.627	44,9	0,8
Negro	347.060	115.064	55,7	0,6	210.506	113.848	55,1	0,8
Faixas de rendimento								
Até 1/4 SM	12.716	3.530	4,1	1,6	8.282	3.641	4,2	1,9
Mais de 1/4 a 1/2 SM	16.091	4.949	5,7	1,3	10.345	5.009	5,8	1,6
Mais de 1/2 a 1 SM	54.539	18.457	21,3	0,8	33.087	17.975	20,7	1,0
Mais de 1 a 2 SM	81.909	33.241	38,3	0,9	51.103	33.083	38,2	1,1
Mais de 2 SM	60.315	26.505	30,6	1,3	39.698	26.981	31,1	1,3
Condição na Força de Trabalho								
<i>Taxa de participação (em %)</i>	459.432	167.535	61,6	0,2	288.672	167.535	61,6	0,3
Na Força de Trabalho	270.229	103.123	61,6	0,6	170.193	103.197	61,6	0,7
Fora da Força de Trabalho	189.203	64.413	38,4	0,5	118.479	64.338	38,4	0,6
Condição de ocupação								
<i>Taxa de desocupação (em %)</i>	270.229	103.123	13,7	0,9	170.193	103.197	13,7	1,1
Ocupado	235.240	88.947	86,3	0,6	149.092	89.045	86,3	0,7
Desocupado	34.989	14.176	13,7	1,1	21.101	14.152	13,7	1,3
Posição na ocupação								
Empregado	150.884	60.516	68,0	0,7	93.828	60.029	67,4	0,8
Empregador	9.914	4.128	4,6	2,1	6.726	4.285	4,8	2,3
Conta própria	64.938	22.112	24,9	0,8	42.071	22.438	25,2	0,9
Trabalhador familiar	9.504	2.190	2,5	2,3	6.467	2.294	2,6	2,6
Grupamento ocupacional								
Diretores e gerentes	10.236	4.512	5,1	2,1	6.863	4.676	5,3	2,4
Prof. das ciências e intelectuais	20.986	9.300	10,5	1,8	13.705	9.398	10,6	1,8
Téc e prof de nível médio	16.887	7.232	8,1	1,6	10.836	7.356	8,3	1,5
Trab de apoio administrativo	15.826	6.744	7,6	1,3	9.914	6.867	7,7	1,6
Trab dos serviços, vendedores	48.488	19.578	22,0	0,9	30.309	19.676	22,1	1,1
Trab qualificados da agropecuária	26.127	5.448	6,1	1,6	17.538	5.553	6,2	1,8
Trab qualificados operários	30.055	12.125	13,6	1	18.701	12.028	13,5	1,3
Operadores e montadores	18.307	7.266	8,2	1,3	11.567	7.248	8,1	1,6
Ocupações elementares	46.149	15.913	17,9	1	28.347	15.429	17,3	1,2
Membros das forças armadas	2.152	817	0,9	4,4	1.292	796	0,9	4,0
Ocupações maldefinidas	27	12	0,0	29,4	20	18	0,0	30,0

Fonte: IBGE. PNAD Contínua. Elaboração própria

Referências bibliográficas

- AMORIM, B.; CORSEUIL, A. H. L. **Análise da dinâmica do emprego setorial de 2014 a 2015**. Nota Técnica n. 23, Brasília: IPEA, 2016.
- BANCO CENTRAL DO BRASIL (BCB). **Relatório de inflação**. Set./2017, Brasília: 2017, pp. 28-30. Disponível em: <
<http://www.bcb.gov.br/htms/relinf/port/2016/09/ri201609b3p.pdf>>.
- FÁVERO, L. *et al.* **Análise de dados**: modelagem multivariada para tomada de decisões. Rio de Janeiro: Elsevier, 2009.
- FREITAS, M.; ANTONACI, G. Sistema Integrado de Pesquisas Domiciliares - Amostra Mestra 2010 e Amostra da Pnad Contínua. **Texto para Discussão n.º 50**, Rio de Janeiro: Diretoria de Pesquisa - IBGE, 2014.
- IBGE. Pesquisa Nacional por Amostra de Domicílio Contínua. **Notas Metodológicas**. Rio de Janeiro: IBGE, 2014.
- _____. Pesquisa Nacional por Amostra de Domicílio Contínua: **Mobilidade sócio-ocupacional**. Rio de Janeiro: IBGE, 2016.
- _____. Pesquisa Nacional por Amostra de Domicílio Contínua. **Notas técnicas – versão 1.4**. Rio de Janeiro: IBGE, 2017.
- MOREIRA, A.; CORSEUIL, C.; FOGUEL, M. **Um modelo de transição entre estados da força de trabalho para previsão do desemprego e da ocupação formal**. Jul.-set./2016, Brasília: IPEA, 2016 (Carta de Conjuntura).
- PIZZOTT, M.; VEIGA, A.; SILVA, D. Indicadores para monitoramento de representatividade no caso da Pesquisa Nacional por Amostra de Domicílios Contínua. *In: IV Workshop Internacional sobre Pesquisas para Avaliação de Políticas Públicas*, 2017, Cuiabá – MT.
- REIS,. Como as condições do mercado de trabalho influenciam as transições do desemprego para o emprego? *In: Mercado de trabalho*. Brasília: IPEA, ano 23, out., 2017.
- RUIZ, C.; SILVA, P. Explorando alternativas para a calibração dos pesos amostrais da Pesquisa Nacional por Amostra de Domicílios. *In: VI Congreso de la Asociación Latinoamericana de Población*, 2014, Lima-Peru.
- TEIXEIRA JUNIOR, A.; ROSSETI, E. S.; ALMEIDA, P. A.; SILVA, D. B. N. Pesos longitudinais para a pesquisa nacional por amostra de domicílios contínua (PNAD Contínua). *Boletim Mercado de Trabalho*, n. 67, set. 2019, Brasília: IPEA, 2019.
- TEIXEIRA JUNIOR, A. A importância do uso de pesos longitudinais na produção de indicadores para análise da evolução do mercado de trabalho. *In: XX Encontro Nacional de Estudos Populacionais*, 2016, Foz do Iguaçu - PR.