

## Revisão da série de rendimentos do trabalho da PNAD Contínua: Identificação e tratamento de rendimentos *outliers* na série 2012/2019 e revisão de valores repetidos na base de 2014

Esta nota técnica apresenta duas revisões metodológicas que o IBGE realizou na base de dados de rendimentos do trabalho da PNAD Contínua. A primeira consiste na identificação e tratamento dos rendimentos do trabalho extremamente elevados (*outliers*) na série disponível, desde janeiro de 2012. A segunda revisão versa sobre a imputação de alguns valores de rendimento das bases do segundo, terceiro e quarto trimestres do ano de 2014.

### 1. Tratamento para outlier nas bases da PNAD Contínua

A incorporação do tratamento dispensado aos rendimentos *outliers* foi motivado por um caso ocorrido na série de 2016. Na ocasião verificou-se um valor muito destoante de rendimento, cuja primeira entrevista foi realizada no quarto trimestre de 2016, que gerou impactos artificiais no aumento do rendimento e da desigualdade em 2016, e, posteriormente, na queda da desigualdade, em 2017. A observação, já objeto de outras notas técnicas<sup>1</sup> foi de um informante, empregador, do setor de transportes, que declarou receber R\$1 milhão de rendimento mensal do trabalho principal – habitual e efetivo.

Visando contornar os impactos causados pelo outlier, para a publicação da Síntese de Indicadores Sociais 2018 do IBGE ficou definido que se utilizaria algum tratamento que amenizasse o efeito deste caso pontualmente. Os indicadores calculados sem o ajuste foram avaliados pelo corpo técnico da publicação, que interpretou que os indicadores de rendimento e de desigualdade apresentavam resultados não compatíveis com a realidade econômica do país. Paralelamente usuários externos também passaram a realizar procedimentos para o tratamento dos rendimentos *outliers*, conforme divulgado em publicações especializadas.<sup>2</sup> Decidiu-se assim realizar um tratamento específico, reduzindo-se o peso desta resposta, conforme descrito na publicação.<sup>3</sup> Na ocasião o alvo foi este único outlier identificado porque causava uma mudança marcante de tendência em indicadores nacionais.

Diante destas ocorrências, o IBGE percebeu a necessidade de criar um procedimento geral para casos em que a resposta de rendimento fosse considerada outlier. Tal procedimento passa a ser aplicado mensalmente nas bases de dados da PNAD Contínua, considerando os bancos de dados com os trimestres móveis. Avaliou-se oportuno também realizar uma revisão com a aplicação do procedimento retroativamente desde o início da série, ou seja, desde janeiro de 2012.

O procedimento metodológico para o tratamento dos *outliers* dividiu-se em duas etapas. Primeiramente a identificação dos valores e posteriormente o tratamento das respostas definidas como outlier – detecção e tratamento.

---

<sup>1</sup> Pesquisa Nacional por Amostra de Domicílios Contínua – PNAD Contínua; Nota Técnica 01/2017. Atualização de valor de rendimento. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9171-pesquisa-nacional-por-amostra-de-domicilios-continua-mensal.html?=&t=notas-tecnicas>

<sup>2</sup> Como por exemplo as matérias divulgadas pelo jornal Valor Econômico, disponíveis em: <https://www.valor.com.br/brasil/5140838/estudo-ve-distorcao-no-aumento-da-renda/>; <https://www.valor.com.br/brasil/5241339/apos-distorcao-ibge-pretende-tratar-outliers-de-pesquisa-domiciliar>; <https://www.valor.com.br/brasil/5491723/milionario-provoca-distorcao-em-pesquisa-de-renda-do-ibge>.

<sup>3</sup> IBGE, Síntese de Indicadores Sociais – 2018, p. 131. Nota técnica 5. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101629>

Seguindo as possibilidades apontadas por Mukherjee et al. (1998)<sup>4</sup> e Filzmoser et al. (2016)<sup>5</sup>, foram inicialmente considerados 3 critérios de identificação de outlier aplicados em duas variáveis: rendimento do trabalho e rendimento total domiciliar per capita, cujos limites seriam definidos em função da média, da mediana ou do terceiro quartil.

Verificou-se que a observação relevante para definição de rendimento outlier é a renda do trabalho, ou seja, corrigindo-se a variável de renda do trabalho, a renda domiciliar per capita estaria preservada. Ao todo foram testadas 87 bases de trimestre móvel, sendo cada base com o tamanho médio de 211 mil domicílios e aproximadamente 565 mil respostas. O princípio norteador da metodologia foi detectar as respostas que representassem apenas casos extremos na distribuição dos rendimentos. Os três critérios de identificação foram testados em nível nacional, por Unidades da Federação (UF) e por Grandes Regiões para os rendimentos do trabalho e para o rendimento domiciliar per capita.

Assim, com os limites superiores definidos em função do rendimento do trabalho principal, habitual e efetivo, detecta-se os valores outliers nas séries de rendimento mensal do trabalho principal e dos demais trabalhos. Os limites têm como propriedade serem flexíveis, tendo assim a capacidade de adaptação aos valores encontrados na distribuição, ou seja, ao contexto econômico – inflação, alta ou baixa incidência de valores altos, por exemplo.

Mais especificamente, os limites superiores foram definidos a partir de três conceitos, recomendados pela literatura mencionada e testados nas distribuições de rendimentos em escala logarítmica:

1.  $m+6dp$  (onde:  $m$  = média;  $dp$  = desvio padrão);
2.  $med+6dm$  (onde:  $med$  = mediana;  $dm$  = desvio mediano absoluto/0,675);
3.  $Q3+6IQ$  (onde:  $Q3$  = terceiro quartil;  $IQ$  = intervalo interquartil);

Avaliou-se que os melhores resultados encontrados nos testes foram com o conceito 1, com os limites definidos por grandes regiões e utilizando o banco trimestral com trimestre móvel.

Uma vez identificados os valores outliers realizou-se o tratamento dos dados. Para esta etapa duas opções metodológicas indicadas internacionalmente foram consideradas: a redução de peso da resposta ou a imputação de valor. Especificamente para pesquisas de força de trabalho a imputação de valor foi mais comumente encontrada.<sup>6</sup> A vantagem desta opção é que os impactos das demais respostas do informante não são alterados, uma vez que seu peso é preservado e somente a resposta de rendimento outlier é imputada.

O tipo de imputação escolhido pelo IBGE para o tratamento dos rendimentos outliers foi determinística. Devido à sua natureza peculiar, ser um rendimento extremamente alto, optou-se por imputar o valor mais alto válido na distribuição, ou seja, que estivesse dentro dos limites estabelecidos pelo método de identificação.

A Tabela 1, a seguir, mostra a quantidade de casos de outliers de rendimentos que foram tratados segundo a metodologia descrita nesta nota. Considerou-se os bancos dos trimestres móveis, que contém os rendimentos habituais e efetivos do trabalho principal e dos demais trabalhos e as cinco entrevistas dos informantes à pesquisa.

---

<sup>4</sup> Mukherjee, C.; White, H.; Wuyts, M. *Econometrics and data analysis for developing countries* (1998).

<sup>5</sup> Filzmoser, Peter, and Matthias Templ. "Detecting outliers in household consumption survey data" (2016).

<sup>6</sup> Como por exemplo no Canadá - Statistics Canada – Canadian Labour Force Survey 2017.

**Tabela 1 – Número de casos identificados e tratados como outlier por ano segundo o tipo de rendimento**

ano	RTH (habitual)		RTE (efetivo)	
	principal	outros	principal	outros
2012	5	1	2	
2013	4		6	1
2014	6		8	
2015	7		9	
2016	6	1	4	
2017	11		13	
2018	1		4	
2019	4		0	1
<b>total</b>	<b>44</b>	<b>2</b>	<b>46</b>	<b>2</b>

Fonte: PNAD Contínua, 2012 a 2019.

Considerando-se apenas as respostas advindas das primeiras entrevistas – que são as utilizadas para os cálculos de indicadores de rendimento médio e de coeficientes de distribuição de renda anuais, as respostas detectadas como outlier e tratadas segundo a descrição acima alcançou o número de 8 para os rendimentos habituais do trabalho e de 10 para os rendimentos efetivos, em todo o período analisado. As Tabelas 2 e 3, a seguir apresentam os dados encontrados como resposta, como limite e como resultado da imputação para ambos os rendimentos – habitual e efetivo.

**Tabela 2 – Rendimentos do trabalho habitual (RTH) para 1ª entrevista – valores originais, limites e valores imputados segundo o tratamento de outlier**

Ano (UF)	(R\$ mil)		
	RTH orig	Limite	RTH imp
2012 (53)	200,0	186,4	50,0
2014 (35)	200,0	151,3	70,0
2014 (31)	200,0	154,9	80,0
2014 (52)	200,0	172,9	40,0
2015 (35)	300,0	191,4	150,0
2016 (35)	1.000,0	218,7	200,0
2017 (53)	300,0	272,2	70,0
2019 (35)	350,0	268,6	200,0

Fonte: PNAD Contínua, 2012 a 2019.

**Tabela 2 – Rendimentos do trabalho efetivo (RTE) para 1ª entrevista – valores originais, limites e valores imputados segundo o tratamento de outlier**

(R\$ mil)

Ano (UF)	RTE orig	Limite	RTE imp
2013 (42)	250,0	155,2	50,0
2014 (35)	200,0	168,4	140,0
2014 (31)	200,0	156,2	80,0
2014 (52)	200,0	178,5	40,0
2015 (35)	300,0	221,8	150,0
2016 (35)	1.000,0	253,9	200,0
2017 (35)	500,0	307,2	300,0
2017 (52)	700,0	242,6	90,0
2018 (41)	450,0	197,4	100,0
2018 (31)	586,0	298,5	100,0

Fonte: PNAD Contínua, 2012 a 2019.

## 2. Tratamento de respostas com valor R\$99 nas bases da PNAD Contínua de 2014

Entre os meses de maio a novembro de 2014 foram registrados 10.776 casos de rendimento do trabalho (trabalho principal, demais trabalhos, rendimento habitual e efetivo) na base da PNAD Contínua com o valor R\$99. Tais casos destoaram do padrão histórico da pesquisa sugerindo a necessidade de tratamento desses valores. Com a possibilidade que esta revisão na série de dados de rendimentos proporciona decidiu-se pelo tratamento de imputação com o sistema regular da PNAD Contínua, que utiliza informações das entrevistas imediatamente anteriores e, em um segundo momento, o *Canadian Census Edit and Imputation System - Canceis*<sup>7</sup>.

<sup>7</sup> CANCEIS user's guide: Canadian census edit and imputation system. Version 4.5. Ottawa: Statistics Canada, Social Survey Methods Division, 2007.