

RADAR IDHM

NOTA METODOLÓGICA RENDA



ipea Instituto de Pesquisa
Econômica Aplicada


FUNDAÇÃO JOÃO PINHEIRO
GOVERNO DE MINAS GERAIS



*Empoderando vidas.
Fortalecendo nações.*

Nota sobre a metodologia de cálculo da renda domiciliar *per capita* a partir dos microdados das PNADs de 2009 a 2014

RESUMO

O objetivo deste trabalho é explicar e documentar a opção adotada pelo projeto *Atlas do Desenvolvimento Humano no Brasil*¹ para o cômputo das estimativas que envolvem a renda domiciliar *per capita* da Pesquisa Nacional por Amostra de Domicílios (PNAD) anual. Para isso, esta nota apresenta alguns estudos e análises realizadas com o intuito de fornecer insumos e alternativas aos pesquisadores e, mais especificamente, esclarecer a opção de tratamento adotada pelo projeto em questão. Ao todo, foram testados e aprofundados cinco métodos de imputação de dados para as seguintes variáveis e indicadores derivados da informação de rendimentos: (i) Média da Renda Domiciliar *Per Capita*; (ii) Índice de Gini; (iii) Taxa de Extrema Pobreza; e, (iv) Dimensão de Renda que compõe o Índice de Desenvolvimento Humano (IDHM) brasileiro. Como resultado, visando maior transparência e ganhos de reprodutibilidade, a sugestão é pela ausência de tratamento nos valores nulos e pela exclusão das observações sem declaração no cômputo das variáveis e indicadores derivados das informações de rendimentos.

1 O projeto é resultado da parceria entre Instituto de Pesquisa Econômica Aplicada (IPEA), Programa das Nações Unidas para o Desenvolvimento (PNUD) e Fundação João Pinheiro (FJP).

INTRODUÇÃO

Dentre os erros não amostrais a que uma pesquisa de coleta de dados está sujeita, a não resposta é um dos que mais se repetem, em maior ou menor grau. A questão que aqui se coloca é: vale a pena corrigir o problema das rendas ignoradas e rendas nulas na Pesquisa Nacional por Amostra de Domicílios (PNAD) anual mediante uma das muitas técnicas disponíveis para a imputação de dados?

As metodologias de imputação para a resolução do problema de dados faltantes nas diversas pesquisas, são desenvolvidas sobre diferentes propostas de abordagens. Deste modo, não existe uma regra única e consensual para tratar do problema quando se refere à informação de renda ignorada ou nula.

Para tentar responder à questão levantada, esta Nota Técnica tem como objetivo dar a conhecer os estudos realizados pelo projeto para selecionar a opção mais adequada no tratamento da renda domiciliar *per capita*. Dessa maneira, a Nota detalha os procedimentos de cinco métodos usados para tratar dados faltantes e nulos de rendimentos, com o intuito de fornecer insumos e alternativas para os pesquisadores e, mais especificamente, esclarecer a opção adotada pelo projeto do *Atlas de Desenvolvimento Humano no Brasil*.

A renda domiciliar *per capita* corresponde a um conjunto de variáveis com algumas peculiaridades que precisam ser explicitadas. As características de trabalho e rendimentos são informações que podem ser prestadas pela própria pessoa, ou por outro morador da mesma unidade familiar, ou então, em último caso, por um não morador da unidade domiciliar.

Quando o agente de pesquisa da PNAD não consegue registrar o valor de uma renda, por exemplo, de uma pessoa que trabalha, mas cuja renda é ignorada por um outro morador que respondeu a entrevista, o campo de resposta é marcado com um código especial referente à renda ignorada (999.999.999.999). E quando uma, dentre as quatorze questões que compõem a renda de apenas um dos membros do domicílio, é ignorada, a renda total deste domicílio é inviabilizada e todos os demais residentes passam a ser marcados com o código especial referente às rendas ignoradas (999.999.999.999).

Uma segunda peculiaridade é que a captação da renda do trabalho se refere ao rendimento mensal médio naquele ano, no trabalho da semana de referência da pesquisa. A referência é limitada ao mês específico da realização da pesquisa (setembro). Com isso, alguns domicílios não têm renda declarada (renda nula) porque não havia ninguém trabalhando na semana de referência da pesquisa. E embora estes domicílios acabem caracterizados como pobres, não o são, ao se levar em consideração as demais características socioeconômicas registradas pela mesma pesquisa. O mais provável é que se trate de um erro de declaração ou coleta, conforme alertam Osório, Soares e Souza (2011) ao demonstrarem que boa parte dos indivíduos com renda zero nas PNADs tem um perfil tão discrepante que apenas é possível explicar sua situação em termos de choques adversos temporários ou de algum tipo de erro (Osório, Soares e Souza; 2011).

As metodologias de imputação aqui apresentadas dizem respeito, portanto, a esses dois pontos principais: (i) indivíduos com rendimentos ignorados – que não quiseram ou não souberam declarar a renda em alguma das fontes (999.999.999.999); e (ii) indivíduos com renda domiciliar *per capita* nula ou igual a zero (R\$0,00).

Para ilustrar como estas duas situações oscilam na PNAD anual ao longo do tempo (Gráfico 1), entre 2004 e 2014, a quantidade de pessoas em domicílio com renda ignorada flutua e apresenta percentual mais elevado nos últimos anos. Em 2013, por exemplo, aproximadamente 10,9 milhões de pessoas residiam em domicílios com renda ignorada, ou, 5,4% da população. Em contrapartida, os rendimentos nulos mantêm-se próximos a 1% ao longo dos anos, com exceção de 2014, ano cujo total de rendas nulas foi consideravelmente reduzido para 0,45%.

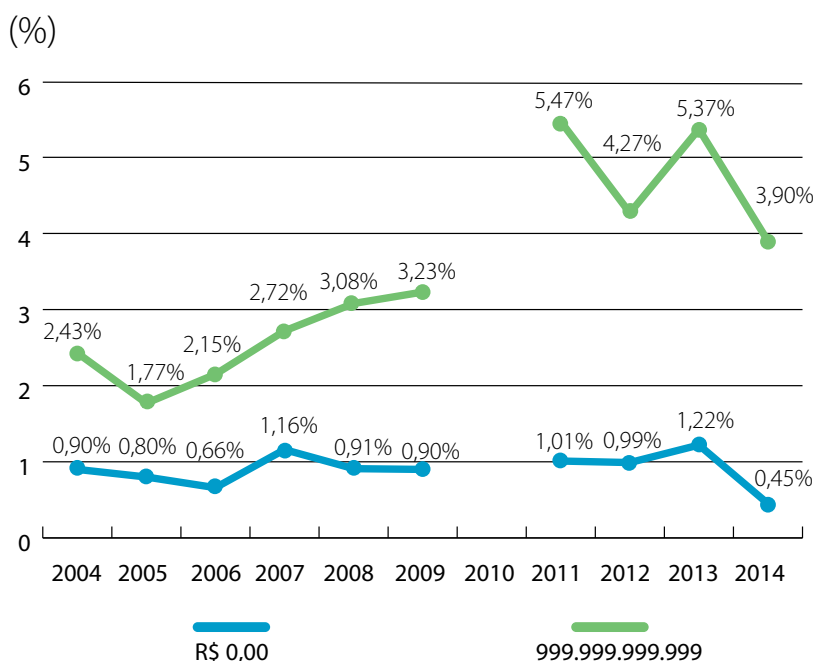


GRÁFICO 1
EVOLUÇÃO
DO TOTAL DE
INDIVÍDUOS COM
RENDA IGNORADA
OU SEM RENDA NA
RENDA DOMICILIAR
PER CAPITA DA
PNAD

Fonte: Pesquisa Nacional por Amostragem Domiciliar (PNAD), 2004 a 2014.
Elaborado pelos autores (2016).

Considerando estas duas situações, é possível supor que a identificação da não declaração é adequada para o caso de renda ignorada (999.999.999.999), mas o mesmo não acontece para as rendas nulas, ou zeros. No montante de rendas nulas, algumas são zeros verdadeiros e outras não, dado que a informação se refere ao rendimento mensal no trabalho da semana de referência.

Vale destacar que, previamente à análise, conforme destaca Rubin (1987), o padrão observado pela não resposta de uma determinada variável é um importante pressuposto para a escolha da metodologia a ser adotada para o tratamento dos dados. Segundo o autor, o padrão de não resposta pode ser classificado em três categorias:

1. *Missing completely at random (MCAR)*: O padrão de não resposta em uma variável Y não depende do valor de Y e nem de demais variáveis X;
2. *Missing at random (MAR)*: O padrão de não resposta em uma variável Y não depende do valor de Y, mas depende de variáveis X;
3. *Missing not completely at random (NMCAR)*: O padrão de não resposta em uma variável Y depende do valor de Y e de variáveis X.

No primeiro caso (MCAR), os dados faltantes de uma variável estudada (Y) são completamente ao acaso e não estão relacionados com nenhuma outra variável. No segundo caso (MAR), o padrão de não resposta não é aleatório e pode ser explicado por variáveis auxiliares (X). E no terceiro caso (NMAR), o dado faltante está relacionado a determinado(s) valor(es) de resposta(s), e pode ser explicado por variáveis auxiliares (X).

Para avançar no tratamento das não respostas das variáveis de rendimentos coletadas pela PNAD anual, pressupõe-se que, tanto o padrão das não respostas, quanto das rendas nulas, são melhores explicados pelo *missing completely at random* (MCAR).

Feita esta observação, diversas são as alternativas possíveis para a imputação de dados, que se dividem em imputação simples ou única, e imputação múltipla. A imputação simples ocorre quando os dados perdidos são substituídos uma única vez por algum dado, dentre os métodos possíveis². Já a imputação múltipla ocorre quando, para cada dado faltante, são imputados vários valores ao invés de um e, posteriormente, obtém-se uma única estimativa através das múltiplas imputações³. Quatro dos métodos aqui testados são de imputação simples e um de imputação múltipla.

Tendo em vista que as rendas nulas são um problema mais complexo, visto que algumas pessoas de fato possuem renda nula, e que o universo das pessoas que declararam possuir renda nula é bastante inferior em relação aos sem declaração, optou-se por trabalhar quatro métodos de imputação de renda para os casos de renda ignorada e apenas um método com tratamento para os casos de renda ignorada e de renda nula, a saber: i) exclusão dos dados; ii) *hot deck* probabilístico; iii) *hot deck* determinístico; iv) imputação múltipla método da média preditiva; e (v) análise discriminante e vizinho mais próximo⁴.

O primeiro método testado, exclusão dos dados, é o mais usual e antigo para tratar o problema da não resposta. Consiste em deletar da base de dados todas as observações que contenham valores faltantes. Entretanto, ao excluir os registros da base de dados, as demais informações do questionário também são deletadas e com isso o tamanho amostral da pesquisa é reduzido. Apesar da facilidade de implementação metodológica e também da ausência de estimadores complexos, ao excluir estes dados, implicitamente se está realizando uma imputação pela média, e isso pode induzir a um viés caso o subgrupo de não-respondentes apresente características distintas das observações restantes na base de dados.

O segundo e terceiro métodos testados, *hot deck* probabilístico e *hot deck* determinístico, derivam do modelo *hot deck* que substitui os dados faltantes por dados de um doador que pode ser um doador único e aleatório (*hot deck* probabilístico), ou então o resultado de uma

.....
2 Dentre os métodos de imputação simples, temos: imputação dedutiva, substituição por um valor de tendência central, imputação geral aleatória, imputação pela média dentro de classes, imputação aleatória dentro de classes, *hot deck*, imputação por regressão preditiva, imputação por regressão aleatória, imputação pela função distância, estimativa de máxima verossimilhança e métodos de imputação única para dados longitudinais.

3 Dentre os métodos de imputação múltipla para padrão monotônico temos: método da regressão linear Bayesiana (BLR- *Bayesian Linear Regression*), método da média preditiva (PMM-*Predictive Mean Matching*). Já para padrão não monotônico temos o MCMC (*Markov Chain Monte Carlo*), o método de Monte Carlo baseado em Cadeia de Markov, que tem como objetivo simular distribuições multivariadas cujo limite é uma cadeia de Markov estacionária que tem a distribuição que se deseja encontrar. É a metodologia adotada por Rubin (1987).

4 Mesmo método de imputação utilizado nos dados do Atlas do Desenvolvimento Humano derivados do Censo Demográfico.

medida estatística descritiva do estrato ou agrupamento de possíveis doadores (*hot deck* determinístico). Em ambos os testes, foram usadas as mesmas variáveis de controle⁵.

A quarta técnica de imputação testada, imputação múltipla via método da média preditiva, recentemente teve o seu uso ampliado em decorrência dos desenvolvimentos computacionais que facilitaram a sua implementação. O método da imputação múltipla via média preditiva combina elementos de regressão, vizinho mais próximo e *hot deck* (Moura *et al.*, 2014).

Por fim, na última técnica, análise discriminante e vizinho mais próximo, são imputados valores de renda domiciliar *per capita* para os casos de renda nula para a população identificada como sem perfil de extrema pobreza, e para todos os que não declararam os rendimentos. Este procedimento estatístico envolveu três passos: i) a aplicação de um “filtro” para selecionar aqueles domicílios que, apesar de aparecerem com renda nula, não possuíam características de extrema pobreza; ii) aplicação de um modelo de análise multivariada para classificar os domicílios; e iii) a atribuição de uma renda domiciliar *per capita* para os renda nula considerados não-extremamente pobres, através da técnica do “vizinho mais próximo”.

Feito este detalhamento das metodologias testadas, avalia-se nos próximos tópicos os resultados na (i) média da renda domiciliar *per capita*, (ii) taxa de extrema pobreza, (iii) índice de gini, e (iv) dimensão de renda (IDHM-R), que compõe o cômputo do IDHM.

RENDIMENTOS MÉDIOS

A renda domiciliar *per capita* se refere à somatória dos rendimentos de todos os moradores de um mesmo domicílio, dividido pelo número total de membros residentes em um mesmo domicílio. Independente do método de imputação de dados utilizado (Tabela 1) no período de 2009 a 2014, a média da renda domiciliar *per capita* cresceu a cada ano.

TABELA 1

COMPARAÇÃO DOS MÉTODOS DE IMPUTAÇÃO DE RENDA PARA A MÉDIA DA RENDA DOMICILIAR *PER CAPITA* DAS PNADS, 2009 A 2014

	Exclusão	<i>Hot Deck</i> Probabilístico	<i>Hot Deck</i> Determinístico	Imputação Múltipla	Discriminante e Vizinho mais Próximo
2009	R\$ 656,97	R\$ 650,19	R\$ 660,59	R\$ 663,71	R\$ 690,75
2011	R\$ 698,48	R\$ 685,10	R\$ 705,47	R\$ 708,25	R\$ 748,64
2012	R\$ 751,67	R\$ 736,93	R\$ 754,44	R\$ 757,06	R\$ 789,55
2013	R\$ 777,56	R\$ 760,58	R\$ 781,29	R\$ 785,44	R\$ 825,01
2014	R\$ 803,36	R\$ 791,43	R\$ 806,22	R\$ 808,55	R\$ 832,31

Fonte: Pesquisa Nacional por Amostragem Domiciliar (PNAD), 2004 a 2014. Elaborado pelos autores (2016).

Nota: Valores deflacionados para agosto de 2010, data de referência do Atlas de Desenvolvimento Humano, segundo Texto para Discussão n. 897 (IPEA, 2002).

De modo geral, o que se observa é que as diferenças entre os métodos de imputação testados, com exceção da análise de discriminante e vizinho mais próximo, não produzem resultados muito divergentes. As diferenças entre os valores extremos das médias da renda domiciliar *per capita* do país para os resultados dos modelos são de R\$40,88, entre o maior valor (análise de discriminante e vizinho mais próximo) e o menor valor (*hot deck*).

ÍNDICE DE GINI

Os métodos de imputação de rendimentos podem afetar de diferentes formas as medidas de desigualdade social. E ao avaliar os efeitos dos cinco métodos de imputação testados para o Índice de Gini (Tabela 2), percebe-se que tanto para a exclusão dos valores *missings* quanto para as três técnicas de imputação há uma queda progressiva do índice nos anos considerados.

TABELA 2

COMPARAÇÃO DOS MÉTODOS DE IMPUTAÇÃO DE RENDA NO ÍNDICE DE GINI DAS PNADS, 2009 A 2014

	Exclusão	<i>Hot Deck</i> Probabilístico	<i>Hot Deck</i> Determinístico	Imputação Múltipla	Discriminante e Vizinho mais Próximo
2009	0,618	0,624	0,619	0,620	0,540
2011	0,606	0,616	0,607	0,608	0,529
2012	0,602	0,611	0,604	0,605	0,526
2013	0,599	0,608	0,600	0,601	0,525
2014	0,591	0,596	0,591	0,592	0,515

Fonte: Pesquisa Nacional por Amostragem Domiciliar (PNAD), 2004 a 2014. Elaborado pelos autores (2016).

Nota: Valores deflacionados para agosto de 2010, data de referência do Atlas de Desenvolvimento Humano, segundo Texto para Discussão n. 897 (IPEA, 2002).

Novamente, a metodologia da análise de discriminante e vizinho mais próximo acaba por impactar a concentração de renda medida pelo Índice de Gini em uma proporção maior do que os outros métodos, que apresentam diferenças pequenas entre eles.

AS TAXAS DE EXTREMA POBREZA

Várias são as abordagens para se definir uma linha de pobreza. Conforme atesta Soares (2009), existem várias métricas que podem ser adotadas e o Brasil não conta com uma linha ou metodologia oficial de pobreza de modo que as políticas públicas e as instituições de pesquisa oficiais utilizam critérios diferentes e contraditórios entre si (Soares, 2009). Portanto, um questionamento primário seria, quando se trata de análises da extrema pobreza, como é definida a linha de corte entre pobres e não pobres.

Os fatores determinantes na definição são tanto de natureza teórica, na determinação conceitual do que é a pobreza, quanto de natureza técnica, que abarca a fonte de dados que será utilizada, o tratamento de algumas variáveis através da imputação de renda, as datas de referência para a linha de corte, etc.

O projeto do Atlas do Desenvolvimento Humano faz uso dos Censos Demográficos de 1991, 2000 e 2010 como fonte da informação e por isso considerou agosto de 2010 como data de referência. Para os demais anos, aplicou-se o deflator conforme Texto para Discussão n. 897 (IPEA, 2002).

Um problema mais relevante do que os valores *missings* para o cálculo das taxas de extrema pobreza são os domicílios com renda zero. Isso porque, para qualquer linha de pobreza adotada, as pessoas com renda zero serão sempre pobres. E o que se pode observar é que a imputação de renda para os dados *missings* influencia pouco nos resultados das taxas de extrema pobreza nas PNADs (Tabela 3).

TABELA 3
COMPARAÇÃO DOS MÉTODOS DE IMPUTAÇÃO DE RENDA NAS TAXAS DE EXTREMA POBREZA, NAS PNADS, 2004 A 2014

	Exclusão	Hot Deck Probabilístico	Hot Deck Determinístico	Imputação Múltipla	Discriminante e Vizinho mais Próximo
2009	5,2	6,1	5,2	5,2	4,3
2011	4,8	6,6	4,8	4,9	3,7
2012	3,9	5,3	4,0	4,0	2,9
2013	4,3	5,8	4,3	4,3	3,1
2014	3,1	4,0	3,1	3,1	2,5

Fonte: Pesquisa Nacional por Amostragem Domiciliar (PNAD), 2004 a 2014. Elaborado pelos autores (2016).

Nota: Valores deflacionados para agosto de 2010, data de referência do Atlas de Desenvolvimento Humano, segundo Texto para Discussão n. 897 (IPEA, 2002).

E independente do modelo adotado, todos os métodos testados revelam um comportamento decrescente na taxa da extrema pobreza ao longo dos últimos anos. Entretanto, a metodologia de imputação de renda *hot deck* probabilístico apresenta resultados mais elevados para a taxa de extrema pobreza, enquanto a análise de discriminante e vizinho mais próximo apresenta resultado mais baixo. Os demais resultados apresentam diferenças pequenas nas taxas de pobreza.

O INDICADOR DE DESENVOLVIMENTO HUMANO (IDHM)

O projeto *Atlas do Desenvolvimento Humano no Brasil* é responsável por elaborar o Índice de Desenvolvimento Humano (IDHM) dos municípios brasileiros, além de outras territorialidades (PNUD, IPEA, FJP; 2013). O cálculo da dimensão do IDHM Renda (IDHM-R) que compõe o IDHM é realizado da seguinte maneira:

$$\text{IDHM-R} = \frac{[\ln(\text{renda per capita}) - \ln(8,00)]}{[\ln(4.033,00) - \ln(8,00)]}$$

Como a fórmula faz uso do logaritmo da média da renda domiciliar per capita, os resultados são semelhantes aos já observados para a média da renda domiciliar *per capita*, porém suavizados (Tabela 4). Por isso, os métodos de imputação testados não produzem resultados muito divergentes entre si e as diferenças entre os modelos são pequenas, da ordem de 0,08 entre o maior valor (análise de discriminante e vizinho mais próximo) e o menor valor (*hot deck* probabilístico) para o ano de 2014.

TABELA 4

COMPARAÇÃO DOS MÉTODOS DE IMPUTAÇÃO DE RENDA NO ÍNDICE DE DESENVOLVIMENTO HUMANO DE RENDA (IDHM-R) DAS PNADS, 2004 A 2014

	Exclusão	<i>Hot Deck</i> Probabilístico	<i>Hot Deck</i> Determinístico	Imputação Múltipla	Discriminante e Vizinho mais Próximo
2009	0,708	0,707	0,709	0,710	0,716
2011	0,718	0,715	0,720	0,720	0,729
2012	0,730	0,727	0,731	0,731	0,738
2013	0,735	0,732	0,736	0,737	0,745
2014	0,741	0,738	0,741	0,742	0,746

Fonte: Pesquisa Nacional por Amostragem Domiciliar (PNAD), 2004 a 2014. Elaborado pelos autores (2016).

Nota: Valores deflacionados para agosto de 2010, data de referência do Atlas de Desenvolvimento Humano, segundo Texto para Discussão n. 897 (IPEA, 2002).

CONCLUSÃO

Embora à primeira vista a questão da imputação das rendas ignoradas pareça importante, do ponto de vista prático, o tratamento de imputação de renda para os “sem declaração” de rendimentos e “renda nula” na PNAD anual mostram impactos pouco significativos, ao se tomar como referência o propósito do *Atlas do Desenvolvimento Humano no Brasil* (ADH), de análise e diagnóstico da evolução de indicadores que podem refletir as características do desenvolvimento humano no país, de modo a contribuir com a formulação, avaliação e monitoramento de políticas públicas.

Previamente à análise, buscou-se compreender as razões pela qual há a não resposta e a renda nula, tendo em vista a sua importância para orientar o manuseio correto da base de dados restante, e também para a escolha de uma possível metodologia de imputação. Se os dados de não resposta ocorressem de forma sistemática, as análises poderiam gerar resultados tendenciosos. Mas os dados de não resposta foram constatados como *missing completely at random* (MCAR), e a amostra restante ainda se mantém representativa da população.

Entre 2009 e 2014, a distribuição de renda captada pela PNAD anual melhorou a cada ano, a média da renda domiciliar *per capita* cresceu, a desigualdade diminuiu e a pobreza, medida por várias linhas, também caiu. E ao se aplicar os cinco modelos de imputação testados – exclusão, *hot deck* probabilístico, *hot deck* determinístico, imputação múltipla e análise de discriminante e vizinho mais próximo – observou-se que os procedimentos de imputação dificultam a reprodução dos dados além de produzir divergências pouco substantivas em relação às estimativas realizadas a partir da exclusão dos casos de renda ignorada.

Sendo assim, concluiu-se que, em contraposição às perdas de informação, a ausência de tratamento produz maior transparência e ganhos na reprodutibilidade dos dados, fator considerado relevante para os propósitos centrais do projeto ADH. E por isso, para o cálculo dos indicadores do ADH, optou-se pela não imputação de dados nos valores nulos e por deletar das bases as observações de rendimentos que apresentavam valores faltantes para o cômputo das variáveis e indicadores derivados da informação de rendimento nas PNADs anuais.

Referências

CORSEUIL, C. H.; FOGUEL, M. N. **Uma sugestão de deflatores para rendas obtidas a partir de algumas pesquisas domiciliares do IBGE**. Rio de Janeiro: Ipea, 2002 (Texto para Discussão n. 897).

CORTEZ, B. F.; MOREIRA, G.G.; FERNANDES, M.V.M. **Descrição e avaliação do processo de imputação nos quesitos de fecundidade da PNAD 2008**. In: XVII Encontro Nacional de Estudos Populacionais, 2010, Caxambú. XVII Encontro Nacional de Estudos Populacionais.

IBGE, 2007. **Imputação de valores faltantes referentes às variáveis de rendimento do trabalho na Pesquisa Mensal de Emprego**. Abril, 2007.

_____, 2012. **Estudos e tratamento da variável rendimento no Censo Demográfico 2010**. Diretoria de Pesquisas, Coordenação de Métodos e Qualidade, Gerência Técnica do Censo Demográfico. Março, 2012.

MDS. **Metodologia de cômputo das taxas de pobreza e extrema pobreza das PNADs 1992 a 2014**. Estudo Técnico 05/2016. Secretaria de Avaliação e Gestão da Informação (SAGI). Brasília, 2016.

_____. **Abordagens para imputação de rendimento na PNAD: método e resultados**. Estudo Técnico 06/2016. Secretaria de Avaliação e Gestão da Informação (SAGI). Brasília, 2016.

MEDEIROS, M. **Medidas de desigualdade e Pobreza**. Brasília, Ed. UnB, 2012.

MOURA, R. G.; ARAVÉQUIA, J. A.; LOPO, A. B. **Preenchimento de falhas em dados de correlação de Anomalia da altura geopotencial (500 hPa)**. Ciência e Natura, Santa Maria, v. 36 Ed. Especial II, 2014, p.503–509.

OSÓRIO, R. SOARES, S.; SOUZA, P. H. G. F. **Erradicar a pobreza extrema: um objetivo ao alcance do Brasil**. Texto para Discussão, n. 1.619. Brasília: Ipea, 2011.

PNUD; IPEA; FJP. **O Índice de Desenvolvimento Humano Municipal Brasileiro: série Atlas do Desenvolvimento Humano no Brasil 2013**. PNUD. Brasília, 2013.

ROCHA, S. A investigação da renda nas pesquisas domiciliares. **Revista Economia e Sociedade**, Campinas, v. 12, n.2 (21), p. 205-224, jul./dez/ 2003.

RUBIN, D. B. **Multiple Imputation for Nonresponse in Surveys**. New York: John Wiley & Sons, 1987.

SOARES, S. S. D. **Metodologia para estabelecer a linha de pobreza: objetivas, subjetivas, relativas, multidimensionais**. Texto para Discussão n. 1381. Rio de Janeiro: Ipea, 2009.

SOUZA, P, H, G, F. **Uma metodologia para decompor diferenças entre dados administrativos e pesquisas, com aplicação para o Programa Bolsa Família e o Benefício de Prestação Continuada na Pnad**. Texto para Discussão n. 1517. Brasília: Ipea, 2010.