



GOVERNO DO DISTRITO FEDERAL - GDF
SECRETARIA DE ESTADO DE ECONOMIA DO DISTRITO FEDERAL - SEEC
COMPANHIA DE PLANEJAMENTO DO DISTRITO FEDERAL - CODEPLAN
DIRETORIA DE ESTUDOS E PESQUISAS SOCIOECONÔMICAS - DIEPS
GERÊNCIA DE PESQUISAS SOCIOECONÔMICAS - GEREPS

PESQUISA DISTRITAL POR AMOSTRA DE DOMICÍLIOS - PDAD 2021

NOTA EXPLICATIVA

Processo de imputação das respostas de local de nascimento

Brasília-DF, julho de 2022

EQUIPE RESPONSÁVEL

DIRETORIA DE ESTUDOS E PESQUISAS SOCIOECONÔMICAS - DIEPS/Codeplan

GERÊNCIA DE PESQUISAS SOCIOECONÔMICAS – GEREPS/DIEPS/Codeplan

NÚCLEO DE PESQUISAS SOCIOECONÔMICAS - NUSOC/GEREPS/DIEPS/Codeplan

Revisão e copidesque

Eliane Menezes - Ascom

SUMÁRIO

1. INTRODUÇÃO	5
2. CORREÇÃO DAS RESPOSTAS DO LOCAL DE NASCIMENTO (SE DISTRITO FEDERAL OU OUTRO LOCAL) DA PDAD 2021	6

1. INTRODUÇÃO

Esta nota explicativa tem por objetivo descrever o processo de imputação das respostas de local de nascimento (se no Distrito Federal ou outro local) da Pesquisa Distrital por Amostra de Domicílios (PDAD) 2021.

2. CORREÇÃO DAS RESPOSTAS DO LOCAL DE NASCIMENTO (SE DISTRITO FEDERAL OU OUTRO LOCAL) DA PDAD 2021

Durante a coleta da PDAD 2021, constatou-se um aparente excesso de marcações de pessoas que nasceram no Distrito Federal. Enquanto, em 2018, o percentual amostral de nascidos no DF ficou em 52,6%; em 2021, o valor foi de 67,9%, de um total de 83.481 pessoas com informações coletadas. Ao comparar esses resultados com as informações da Pesquisa de Emprego e Desemprego (PED), realizada também pela Codeplan em conjunto com o DIEESE, constatou-se que o número de 2018 era, de fato, o mais consistente.

Dessa forma, ficou evidente uma discrepância de valores, o que implicaria divulgação de informações que não refletiriam a realidade, dada a magnitude da mudança em um período implausível de tempo. Para realizar a correção, optou-se por utilizar conjuntamente a amostra da PDAD 2018 e 2021, uma vez que ambas as pesquisas seguiram metodologias e questionários semelhantes. Para tanto, foram definidas variáveis auxiliares, de modo a se tentar ajustar da melhor maneira possível a chance de a pessoa ter, de fato, nascido no Distrito Federal. As variáveis utilizadas estão descritas no Quadro 1.

Quadro 1 - Descrição das variáveis elencadas para imputação das informações de local de nascimento da PDAD 2021

(Continua)

Variável	Descrição
Região Administrativa	A correção foi realizada para cada uma das 33 Regiões Administrativas.
Pontuação do Critério Brasil	Trata-se de uma metodologia desenvolvida pela Associação Brasileira de Empresas de Pesquisa (ABEP) para mensurar o poder de compra das famílias brasileiras. Diante da falta de informações de renda, que, usualmente, apresenta difícil captação, são utilizadas informações alternativas para divisão das famílias em classes sociais, com níveis de renda associados indiretamente. As informações completas sobre essa metodologia estão disponíveis em: https://www.abep.org/criterioBr/01_cceb_2021.pdf Este critério considera a posse/existência dos seguintes itens no domicílio: • Quantidade de banheiros • Quantidade de empregados domésticos • Quantidade de automóveis • Quantidade de microcomputadores • Quantidade de lava-louças • Quantidade de geladeiras • Quantidade de freezers • Quantidade de lava-roupas • Quantidade de aparelhos reprodutores de DVD • Quantidade de micro-ondas • Quantidade de motocicletas • Quantidade de secadoras de roupas Além disso, considera-se a escolaridade do chefe da família (Analfabeto / Fundamental I incompleto; Fundamental I completo / Fundamental II incompleto; Fundamental II completo / Médio incompleto; Médio completo / Superior incompleto; e Superior completo) e, adicionalmente, é considerado o acesso do domicílio a dois serviços públicos: água encanada e rua pavimentada.

Quadro 1 - Descrição das variáveis elencadas para imputação das informações de local de nascimento da PDAD 2021 (Conclusão)

Variável	Descrição
	Como a PDAD não investiga água encanada, utiliza-se como equivalente direto o acesso à rede geral de abastecimento da CAESB. Para cada uma das respostas é atribuída uma pontuação que, quando somadas, variam entre 0 e 100 pontos. Essa pontuação é utilizada para a correção.
Sexo de nascimento	Utiliza-se o sexo de nascimento do respondente.
Pretos e pardos	É utilizada a composição de pretos e pardos como categoria para a característica de raça/cor da pele.
Cônjuge ou companheiro	É criada uma categoria para pessoas que vivem em companhia de cônjuge ou companheiro. Para pessoas as quais as perguntas não se aplicavam (menores de 14 anos, considerou-se que elas não viviam em companhia de cônjuge ou companheiro).
Estudante	Variável binária para verificar se a pessoa era estudante do ensino regular ou creche.
Ocupado	Variável binária para identificar se a pessoa estava trabalhando. Para pessoas as quais as perguntas não se aplicavam (menores de 14 anos em 2018 e menores de dez anos em 2021), considerou-se que elas não trabalhavam.
Quantidade de Filho	Quantidade de pessoas na posição de filho no domicílio.
Plano de saúde privado	Uma variável binária indicando se a pessoa possuía plano de saúde privado.
Escolaridade	Escolaridade das pessoas, considerando-se as seguintes categorias: sem escolaridade; fundamental incompleto; fundamental completo; médio incompleto; médio completo; superior completo; superior incompleto.

Fonte: GEREPS/DIEPS/Codeplan

Para realizar a correção, fez-se necessário determinar a quantidade de casos que passariam por reclassificação. Todas as pessoas com respostas de nascimento fora do Distrito Federal em 2021 tiveram sua marcação inalterada, uma vez que o problema se referia ao excesso de pessoas nascidas no próprio Distrito Federal. Com isso, para pessoas nascidas no DF, havia um excesso indeterminado de pessoas que, provavelmente, estavam com marcação equivocada.

Na avaliação por Região Administrativa, observou-se que, para algumas delas, a discrepância entre os valores coletados em 2018 e 2021 era mais elevada. Assim sendo, foram determinados três percentuais de casos problemáticos (60%, 65% e 70%) que, aleatoriamente, passariam pelo método de correção. Posteriormente, para cada um desses resultados, selecionou-se aquele que apresentava o menor desvio absoluto em relação aos valores observados em 2018, dando-se, sempre que possível, preferência para diferenças positivas.¹ O percentual de casos selecionados para correção, para cada Região Administrativa, é apresentado na Tabela 1.

Após análise das opções tecnológicas disponíveis para a imputação de dados, fez-se uso do pacote "mice" do software R, desenvolvido por van Buuren e Groothuis-Oudshoorn (2011). O pacote realiza imputações por diversos métodos, tendo como foco a imputação multivariada por cadeia de equações (*Multivariate Imputation by Chained Equations* - MICE).² Para o caso específico do problema aqui identificado, por se tratar de uma variável binária (ter nascido ou não no Distrito Federal), o método aplicado foi uma regressão logística.

¹ A tendência histórica é de um aumento paulatino de pessoas nascidas no DF, uma vez que a maioria das pessoas que iniciou a ocupação da capital federal permaneceu e teve filhos e netos no local.

² Informações sobre o funcionamento do pacote estão disponíveis em <https://www.jstatsoft.org/article/view/v045i03>. A versão utilizada foi a 3.14.0.

Tabela 1 - Percentual e número de casos que passaram por imputação na variável de local de nascimento da PDAD 2021, por Região Administrativa

Região Administrativa	% do total problemático	Nº de observações na RA	% do total de observações na RA
Plano Piloto	60%	6.462	59%
Gama	70%	527	28%
Taguatinga	70%	2.442	55%
Brazlândia	65%	1.274	71%
Sobradinho	60%	899	49%
Planaltina	70%	379	11%
Paranoá	70%	337	21%
Núcleo Bandeirante	65%	485	39%
Ceilândia	70%	758	32%
Guará	65%	1.118	38%
Cruzeiro	65%	230	18%
Samambaia	70%	579	22%
Santa Maria	70%	427	24%
São Sebastião	70%	795	24%
Recanto das Emas	65%	350	20%
Lago Sul	60%	1.534	58%
Riacho Fundo	65%	511	31%
Lago Norte	60%	1.437	60%
Candangolândia	60%	684	51%
Águas Claras	60%	2.222	55%
Riacho Fundo II	70%	580	33%
Sudoeste/Octogonal	70%	835	35%
Varjão	65%	514	36%
Park Way	70%	1.090	54%
SCIA	70%	725	21%
Sobradinho II	70%	946	52%
Jardim Botânico	65%	1.477	38%
Itapoã	65%	687	44%
SIA	65%	251	34%
Vicente Pires	70%	1.513	32%
Fercal	70%	489	34%
Sol Nascente/Pôr do Sol	70%	831	49%
Arniqueira	65%	370	25%
Total		33.758	40%

Fonte: GEREPS/DIEPS/Codeplan

Como se trata de um método multivariado, fez-se necessário determinar covariáveis que ajudassem a determinar o valor ausente. Comumente, toda a base de dados é utilizada para esta tarefa. Todavia, como a PDAD é uma pesquisa muito extensa, a utilização de toda a base de dados inviabilizaria o processo; além disso, a utilização de certas variáveis poderia não fazer sentido teórico. Assim sendo, as variáveis elencadas são aquelas apresentadas anteriormente no Quadro 1.

O resultado da imputação, por Região Administrativa, é apresentado na Tabela 2.

Tabela 2 - Resultado da imputação de pessoas nascidas no Distrito Federal antes e após a imputação e comparação com os resultados de 2018, por Região Administrativa, PDAD 2021

Região Administrativa	2018	2021 (Antes)	2021 (imputada)	Diferença absoluta 2021 (antes)/2018	Diferença absoluta 2021 (imputada)/2018
Plano Piloto	39,3%	72,9%	39,4%	33,6%	0,1%
Gama	61,5%	72,0%	65,6%	10,5%	4,1%
Taguatinga	54,1%	72,2%	54,3%	18,1%	0,3%
Brazlândia	67,4%	85,2%	62,9%	17,8%	4,5%
Sobradinho	52,6%	71,6%	50,1%	19,0%	2,5%
Planaltina	56,9%	61,8%	56,9%	4,9%	0,0%
Paranoá	59,6%	67,8%	60,3%	8,2%	0,7%
Núcleo Bandeirante	47,8%	71,0%	56,5%	23,2%	8,8%
Ceilândia	56,8%	65,3%	54,4%	8,5%	2,4%
Guará	49,6%	67,0%	51,8%	17,4%	2,3%
Cruzeiro	45,5%	54,5%	46,9%	9,0%	1,4%
Samambaia	59,6%	64,8%	57,3%	5,2%	2,3%
Santa Maria	57,7%	66,3%	58,1%	8,6%	0,4%
São Sebastião	43,7%	55,8%	44,9%	12,1%	1,3%
Recanto das Emas	55,8%	61,0%	54,9%	5,2%	0,9%
Lago Sul	44,5%	73,8%	45,8%	29,3%	1,3%
Riacho Fundo	51,5%	61,7%	51,2%	10,3%	0,2%
Lago Norte	45,0%	75,4%	46,9%	30,4%	1,9%
Candangolândia	52,8%	68,0%	45,0%	15,2%	7,8%
Águas Claras	44,0%	73,8%	44,7%	29,8%	0,7%
Riacho Fundo II	60,0%	68,2%	58,5%	8,3%	1,4%
Sudoeste/Octogonal	43,4%	57,8%	38,9%	14,4%	4,5%
Varjão	48,4%	58,0%	44,2%	9,6%	4,2%
Park Way	47,1%	77,6%	54,1%	30,5%	7,0%
SCIA	50,3%	58,6%	50,3%	8,3%	0,0%
Sobradinho II	52,3%	73,0%	55,9%	20,7%	3,6%
Jardim Botânico	53,3%	66,3%	51,7%	13,0%	1,7%
Itapoã	50,7%	70,6%	50,4%	19,9%	0,2%
SIA	28,2%	50,5%	22,5%	22,3%	5,8%
Vicente Pires	54,3%	69,9%	57,5%	15,6%	3,2%
Fercal	58,0%	65,9%	54,5%	7,9%	3,5%
Sol Nascente/Pôr do Sol	53,1%	69,6%	51,2%	16,5%	1,9%
Amiqueira	50,3%	62,7%	53,9%	12,4%	3,6%
Total	52,60%	67,90%	50,40%	15,30%	2,2%

Fonte: GEREPS/DIEPS/Codeplan

Após passar pelo processo de correção, algumas variáveis não tiveram seu valor original alterado, ou seja, as pessoas nascidas no Distrito Federal permaneceram com a mesma marcação. Assim, a Tabela 3 apresenta, por Região Administrativa, os casos que efetivamente tiveram mudança de marcação.

Tabela 3 - Casos que efetivamente tiveram mudança de marcação, por Região Administrativa

Região Administrativa	N	%
Plano Piloto	3.695	34%
Gama	122	7%
Taguatinga	807	18%
Brazlândia	414	23%
Sobradinho	403	22%
Planaltina	179	5%
Paranoá	121	8%
Núcleo Bandeirante	184	15%
Ceilândia	273	11%
Guará	455	16%
Cruzeiro	108	8%
Samambaia	232	9%
Santa Maria	163	9%
São Sebastião	390	12%
Recanto das Emas	118	7%
Lago Sul	754	28%
Riacho Fundo	196	12%
Lago Norte	702	29%
Candangolândia	315	23%
Águas Claras	1.184	29%
Riacho Fundo II	184	10%
Sudoeste/Octogonal	469	20%
Varjão	208	15%
Park Way	478	24%
SCIA	340	10%
Sobradinho II	315	17%
Jardim Botânico	582	15%
Itapoã	316	20%
SIA	208	28%
Vicente Pires	608	13%
Fercal	164	12%
Sol Nascente/Pôr do Sol	312	19%
Arniqueira	142	10%
Total	15.141	18%

Fonte: GEREPS/DIEPS/Codeplan

Verifica-se, assim, que, no geral, cerca de 18% das observações tiveram sua marcação de local de nascimento alterada, o que gerou uma diferença absoluta média de 2,2 pontos percentuais entre os resultados da amostra de 2018 e da amostra de 2021.

**Companhia de Planejamento
do Distrito Federal - Codeplan**

Setor de Administração Municipal
SAM, Bloco H, Setores Complementares
Ed. Sede Codeplan
CEP: 70620-080 - Brasília-DF
Fone: (0xx61) 3342-2222
www.codeplan.df.gov.br
codeplan@codeplan.df.gov.br