

PESQUISA NACIONAL POR AMOSTRA DE DOMICÍLIOS CONTÍNUA (PNADC)

Dia 4 - Análises em R

Thiago Cordeiro Almeida

Doutorando, Centre d'Estudis Demogràfics (CED, Espanha)

Pesquisador Assistente (Cebrap)

October 10, 2025



ANTES DE COMEÇAR...

lista de presença!



ANTES DE COMEÇAR (2)...

- Dúvidas gerais sobre a aula anterior?
- Exercício dos dias anteriores: dúvidas, comentários, considerações?



ESTRUTURA DA AULA

Tópicos que vamos cobrir hoje são:

- Tipos de Medida na PNADC
- Análise prática a partir de um caso recente
- Tópicos complementares
- Próximos passos



TIPOS DE MEDIDA NA PNADC

Análise de dados, pesquisa e R



ANÁLISE DE DADOS E PESQUISA

Fluxo de análise de dados¹ que, geralmente, seguimos:

Importar



Arrumar

(Armazenar os dados consistentemente)



Transformar

(Criar novas variáveis e agregações)

Visualizar

(Surpreende, mas não é escalável)

Modelar

(É escalável, mas não surpreende)

Comunicar

Automatizar



NATUREZA DAS VARIÁVEIS

São categorizações excludentes, *todavia, uma variável pode ser transformada em outra natureza/tipo.*

QUALITATIVA

Valores apresentam atributos ou qualidades.

- **Nominal** (não ordenado) - pode ainda ser dividida em binária ou não.
 - Ex.: estados, sexo, nomes.
- **Ordinal** (ordenado)
 - Ex.: nível de escolaridade, **faixa de renda**, faixa de tamanho de cidade

QUANTITATIVA

Valores são expressos em números, apresentam quantidades.

- **Discreta** (contagem) - expressa em números inteiros.
 - Ex.: anos de estudos, número de irmãos, número de domicílios.
- **Contínua** (medidas) - expressa em números infinitesimais.
 - Ex.: Altura, peso, **renda**



NATUREZA DAS VARIÁVEIS EM R

Como verificar: no R, podemos verificar o seu tipo/natureza usando funções, como “class()” ou “str()”.

QUALITATIVA

Valores apresentam atributos ou qualidades.

- **Nominal** (não ordenado) - pode ainda ser dividida em binária ou não.
 - R: character / factor.
- **Ordinal** (ordenado)
 - R: factor

QUANTITATIVA

Valores são expressos em números, apresentam quantidades.

- **Discreta** (contagem) - expressa em números inteiros.
 - R: Integer.
- **Contínua** (medidas) - expressa em números infinitesimais.
 - R: double



TIPOS DE MEDIDAS

Para cada tipo de variável, há um tipo de medida que se adequa melhor à sua análise.

- **Números absolutos**
- **Números relativos** (distribuição relativa; composição)
- **Proporção** (numerador está contido no denominador)
- **Taxa** (numerador **não** está contido no denominador)
- **Razão** (divisão de duas grandezas quaisquer)
- **Medidas de tendência central** (média, moda, mediana)
- **Medidas de variabilidade/dispersão** (variância, desvio padrão, coeficiente de variação)



TIPOS DE MEDIDAS NO R USANDO PLANO AMOSTRAL

Em destaque abaixo, há uma forma de implementação de cada medida em R.

- **Números absolutos** = `svytotal(~VARIAVEL, design = BASE_COM_PLANO_AMOSTRAL)`
- **Números relativos** (distribuição relativa; composição) = `svymean(~VARIAVEL, design = BASE_COM_PLANO_AMOSTRAL)`
- **Proporção** (numerador está contido no denominador) = `svymean(~VARIAVEL, design = BASE_COM_PLANO_AMOSTRAL)`
- **Taxa** (numerador não está contido no denominador) = `svyratio(~NUMERADOR, ~DENOMINADOR, design = BASE_COM_PLANO_AMOSTRAL)`
- **Razão** (divisão de duas grandezas quaisquer) = `svyratio(~NUMERADOR, ~DENOMINADOR, design = BASE_COM_PLANO_AMOSTRAL)`
- **Medidas de tendência central** (média, moda, mediana) = `svymean(~NUMERADOR, ~DENOMINADOR, design = BASE_COM_PLANO_AMOSTRAL)`, `svyquantile(~VARIAVEL, quantiles=~c(.5), design = BASE_COM_PLANO_AMOSTRAL)`
- **Medidas de variabilidade/dispersão** (variância, desvio padrão, coeficiente de variação) = `confint(tabela)`, `cv(tabela)`, `SE(tabela)`



TIPOS DE ANÁLISES

Há níveis de análises que são necessários de serem explorados na interação de fenômenos/variáveis.

UNIVARIADA

Variável qualitativa
(binária, ordinal)

Variável quantitativa
(contínua, discreta)

BIVARIADA

Quali X Quanti

Quali X Quali

Quanti X Quanti

MULTIVARIADA

Quali X Quali X Quali

Quali X Quali X Quanti

Quanti X Quanti X Quali

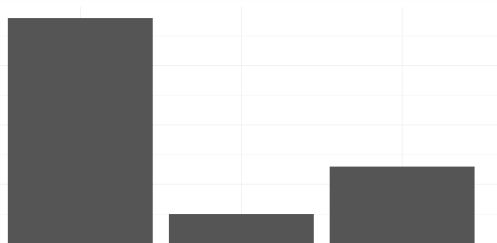
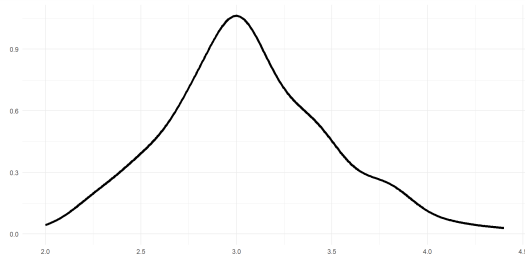
Quanti X Quanti X Quati



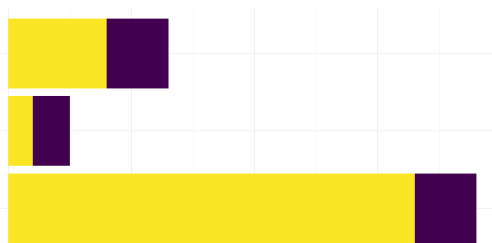
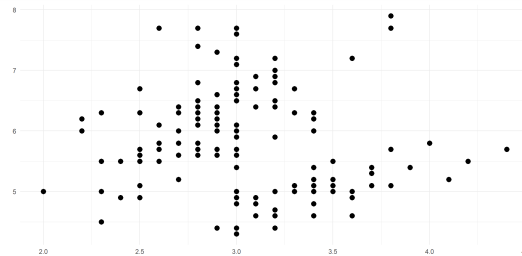
TIPOS DE ANÁLISES

Há níveis de análises que são necessários de serem explorados na interação de fenômenos/variáveis¹.

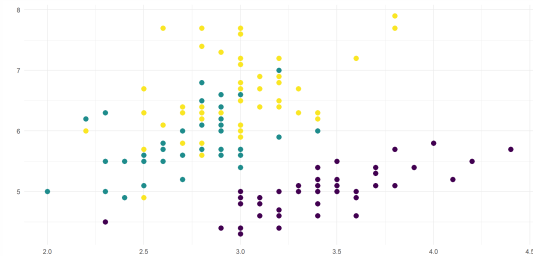
UNIVARIADA



BIVARIADA



MULTIVARIADA



ANÁLISE PRÁTICA A PARTIR DE UM CASO RECENTE

Imposto de renda zero!



O CASO

A ideia da PL que foi para tramitação no Senado Federal de ser um passo rumo à uma política fiscal mais redistributiva.



- ✓ **Quem recebe até R\$ 5.000:** isenção total, economia anual de R\$ 4.356,89.
- ✓ **Quem recebe R\$ 5.500:** a isenção é de 75%, economia anual de R\$ 3.367,68.
- ✓ **Quem recebe R\$ 6.000:** a isenção é de 50%, economia anual de R\$ 2.350,79.
- ✓ **Quem recebe R\$ 6.500:** a isenção é de 25%, economia anual de R\$ 1.333,90.
- ✓ **Quem recebe acima de R\$ 7.000:** As alíquotas progressivas de 7,5%, 15%, 22,5% e 27,5%, permanecem conforme é hoje em dia, sem aumento



O CASO

A ideia da PL que foi para tramitação no Senado Federal de ser um passo rumo à uma política fiscal mais redistributiva.

PROBLEMÁTICA

Qual seria a parcela dos trabalhadores de um específico estado que *potencialmente* seriam beneficiados com imposto zero?

BRASIL MAIS JUSTO
Imposto de Renda zero
para quem ganha até 5 mil

TAREFAS PARA RESPONDER À PROBLEMÁTICA

1. Número de pessoas ocupadas no Brasil.
2. Número de pessoas Ocupadas com potencialmente beneficiados.
3. Proporção de pessoas ocupadas potencialmente beneficiados entre os ocupados.
4. Renda média do grupo dos potencialmente beneficiados ou não.
5. Proporção de potencialmente beneficiados por UF.



O CASO

A ideia da PL que foi para tramitação no Senado Federal de ser um passo rumo à uma política fiscal mais redistributiva.

PROBLEMÁTICA

Qual seria a parcela dos **trabalhadores** de um específico **estado** que *potencialmente* seriam beneficiados com imposto zero?

BRASIL MAIS JUSTO
Imposto de Renda zero
para quem ganha até 5 mil

TAREFAS PARA RESPONDER À PROBLEMÁTICA

1. Número de pessoas ocupadas no Brasil.
2. Número de pessoas Ocupadas com potencialmente beneficiados.
3. Proporção de pessoas ocupadas potencialmente beneficiados entre os ocupados.
4. Renda média do grupo dos potencialmente beneficiados ou não.
5. Proporção de potencialmente beneficiados por UF.



O CASO NA PNADC

Dados do
acumulado
na 1ª visita

Entrevista	Núcleo Básico		Temas	Ciclos de coleta de suplementos			
	SIPD	Pesquisa		1º Trim	2º Trim	3º Trim	4º Trim
1ª Entrevista	Sexo	Trabalho Rendimento	- Características Adicionais do Mercado de Trabalho - Rendimento de outras fontes - Características gerais dos moradores* - Habitação	- Covid-19 (somente em 2023)	- Educação (exceto 2020 e 2021) - Atenção Primária à Saúde (somente 2022)	- Turismo ** - Pessoas com Deficiência (somente em 2022)	- TIC (Exceto em 2020 e 2023) - Sensação de Segurança (somente em 2021) - Furto e roubo (somente em 2021) - Teletrabalho e trabalho por aplicativo (somente em 2022) - Segurança alimentar (somente em 2023)
2ª Entrevista	Cor ou Raça		- Turismo **				
3ª Entrevista	Idade						
4ª Entrevista	Composição do domicílio						
5ª Entrevista	Nível de Instrução		- Rendimento de outras fontes *** - Características gerais dos moradores* - Outras formas de trabalho - Trabalho de crianças e adolescentes				



O CASO NA PNADC

Dados do
acumulado
na 1ª visita



Entrevista	Núcleo Básico		Temas	Ciclos de coleta de suplementos			
	SIPD	Pesquisa		1º Trim	2º Trim	3º Trim	4º Trim
1ª Entrevista	Sexo	Trabalho Rendimento	- Características Adicionais do Mercado de Trabalho - Rendimento de outras fontes - Características gerais dos moradores* - Habitação	- Covid-19 (somente em 2023)	- Educação (exceto 2020 e 2021) - Atenção Primária à Saúde (somente 2022)	- Turismo ** - Pessoas com Deficiência (somente em 2022)	- TIC (Exceto em 2020 e 2023) - Sensação de Segurança (somente em 2021) - Furto e roubo (somente em 2021) - Teletrabalho e trabalho por aplicativo (somente em 2022) - Segurança alimentar (somente em 2023)
2ª Entrevista	Cor ou Raça		- Turismo **				
3ª Entrevista	Idade						
4ª Entrevista	Composição do domicílio						
5ª Entrevista	Nível de Instrução		- Rendimento de outras fontes *** - Características gerais dos moradores* - Outras formas de trabalho - Trabalho de crianças e adolescentes				

Por que **Visita 1** e não a **Visita 5** ou trimestre?



CRIAÇÃO DE VARIÁVEIS

Estado/UF

- Variável da PNADC (UF)

População Ocupada

- Baseado na condição de ocupação (VD4002)

Potencialmente beneficiados

- Com base na renda mensal do trabalho principal (VD4017)
 - Beneficiário: renda menor ou igual a R\$ 5.000,00
 - Não beneficiário: acima de R\$ 5.000,00

Código da variável	Quesito		Categorias		Período
	n°	descrição	Tipo	Descrição	
UF		Unidade da Federação	11	Rondônia	1º tri/2024 - 4º tri/2024
			12	Acre	
			13	Amazonas	
			14	Roraima	
			15	Pará	
			16	Amapá	
			17	Tocantins	
			21	Maranhão	
			22	Piauí	
			23	Ceará	
			24	Rio Grande do Norte	
			25	Paraíba	
			26	Pernambuco	
			27	Alagoas	
			28	Sergipe	
			29	Bahia	
			31	Minas Gerais	
			32	Espírito Santo	
			33	Rio de Janeiro	
			35	São Paulo	
			41	Paraná	
			42	Santa Catarina	
			43	Rio Grande do Sul	
			50	Mato Grosso do Sul	
			51	Mato Grosso	
			52	Goiás	
			53	Distrito Federal	

Código da variável	Quesito		Categorias		Período
	n°	descrição	Tipo	Descrição	
VD4002		Condição de ocupação na semana de referência para pessoas de 14 anos ou mais de idade	1	Pessoas ocupadas	1º tri/2024 - 4º tri/2024
			2	Pessoas desocupadas	
				Não aplicável	

Código da variável	Quesito		Categorias		Período
	n°	descrição	Tipo	Descrição	
VD4017		Rendimento mensal efetivo do trabalho principal para pessoas de 14 anos ou mais de idade (apenas para pessoas que receberam em dinheiro, produtos ou mercadorias no trabalho principal)	Valor	R\$	1º tri/2024 - 4º tri/2024
				Não aplicável	





(PARÊNTESES)

Distintas formas de importação dos dados



FORMAS DE IMPORTAÇÃO DOS DADOS

```
1 Pnad_2022_2s <- get_pnadc(  
2  
3   year = 2022, # obrigatório de ser declarado  
4   quarter/interview/topic = , #  
5   design = TRUE, # opcional, utilizado caso queira aplicar o plano amostral complexo aos dados  
6   vars = c("codigo_variavel1", "codigo_variavel2", "codigo_variavelN"), # opcional  
7   labels = TRUE, # é o default, para rotular os dados  
8   deflator = TRUE # é o default, para incluir variável de deflacionamento direto do IBGE  
9 )
```

O que vai, basicamente variar é o argumento `quarter/interview/topic`.

- Devemos escolher um deles, a depender do nosso interesse de análise.
- Acumulado na entrevista: `interview`;
- Acumulado no trimestre: `quarter`
- Suplemento acumulado no trimestre: `topic`

Lembrem-se que **esses argumentos (quarter/interview/topic) são excludentes no pacote**. Ou seja, deve-se escolher um deles.



FORMAS DE IMPORTAÇÃO DOS DADOS

1. Importação de dados trimestrais:

```

1 Pnad_2022_2s <- get_pnadc(
2
3   year = 2022, # obrigatório de ser declarado
4   quarter = 2, # utilizado somente quando importaremos dados em determinado tri
5   design = TRUE, # opcional, utilizado caso queira aplicar o plano amostral com
6   vars = c("codigo_variavel1", "codigo_variavel2", "codigo_variavelN"), # opcional
7   labels = TRUE, # é o default, para rotular os dados
8   deflator = TRUE # é o default, para incluir variável de deflacionamento direto
9 )

```

Núcleo Básico	
SIPD	Pesquisa
Sexo	Trabalho Rendimento
Cor ou Raça	
Idade	
Composição do domicílio	
Nível de Instrução	



FORMAS DE IMPORTAÇÃO DOS DADOS

2. Importação dos dados ACUMULADOS em determinada visita:

```

1 Pnad_2022_2s <- get_pnadc(
2   year = 2022, # obrigatório de ser declarado
3   interview = 1, # utilizado somente quando importaremos dados em determinada vi.
4   design = TRUE, # opcional, utilizado caso queira aplicar o plano amostral comp
5   vars = c("codigo_variavel1", "codigo_variavel2", "codigo_variavelN"), # opcional
6   labels = TRUE, # é o default, para rotular os dados
7   deflator = TRUE # é o default, para incluir variável de deflacionamento direto
8 )

```

Temas
<ul style="list-style-type: none"> - Características Adicionais do Mercado de Trabalho - Rendimento de outras fontes - Características gerais dos moradores* - Habitação
- Turismo **
<ul style="list-style-type: none"> - Rendimento de outras fontes *** - Características gerais dos moradores* - Outras formas de trabalho - Trabalho de crianças e adolescentes



FORMAS DE IMPORTAÇÃO DOS DADOS

3. Importação de SUPLEMENTO de dados ACUMULADOS em determinado trimestre:

Variáveis iniciadas com *V* ou *VI*: ver sempre o dicionário

```
1 Pnad_2022_2s <- get_pnadc(
2   year = 2022, # obrigatório de ser declarado
3   topic = 4, # utilizado somente quando importaremos dados em determinado trimes
4   design = TRUE, # opcional, utilizado caso queira aplicar o plano amostral comp
5   vars = c("codigo_variavel1", "codigo_variavel2", "codigo_variavelN"), # opciona
6   labels = TRUE, # é o default, para rotular os dados
7   deflator = TRUE # é o default, para incluir variável de deflacionamento direto
8 )
```

Ciclos de coleta de suplementos

1º Trim	2º Trim	3º Trim	4º Trim
- Covid-19 (somente em 2023)	- Educação (exceto 2020 e 2021) - Atenção Primária à Saúde (somente 2022)	- Turismo ** - Pessoas com Deficiência (somente em 2022)	- TIC (Exceto em 2020 e 2023) - Sensação de Segurança (somente em 2021) - Furtos e roubos (somente em 2021) - Teletrabalho e trabalho por aplicativo (somente em 2022) - Segurança alimentar (somente em 2023)



FORMAS DE IMPORTAÇÃO DOS DADOS

4. Importação dos dados ACUMULADOS em determinado trimestre para suplementos específicos:

Variáveis iniciadas com *S*: ver sempre o dicionário

```
1 Pnad_2022_2s <- get_pnadc(
2   year = 2022, # obrigatório de ser declarado
3   topic = 4, # utilizado somente quando importaremos dados em determinado trime.
4   selected = TRUE, # utilizado para suplementos coletados para alguns moradores
5   design = TRUE, # opcional, utilizado caso queira aplicar o plano amostral com
6   vars = c("codigo_variavel1", "codigo_variavel2", "codigo_variavelN"), # opcional
7   labels = TRUE, # é o default, para rotular os dados
8   deflator = TRUE # é o default, para incluir variável de deflacionamento direto
9 )
```

Ciclos de coleta de suplementos

1º Trim	2º Trim	3º Trim	4º Trim
- Covid-19 (somente em 2023)	- Educação (exceto 2020 e 2021) - Atenção Primária à Saúde (somente 2022)	- Turismo ** - Pessoas com Deficiência (somente em 2022)	- TIC (Exceto em 2020 e 2023) - Sensação de Segurança (somente em 2021) - Furto e roubo (somente em 2021) - Teletrabalho e trabalho por aplicativo (somente em 2022) - Segurança alimentar (somente em 2023)



VAMOS PARA O R!

Abram o arquivo: dia 4 >

praticas > codigos > Aula 4 -
roteiro completo.R



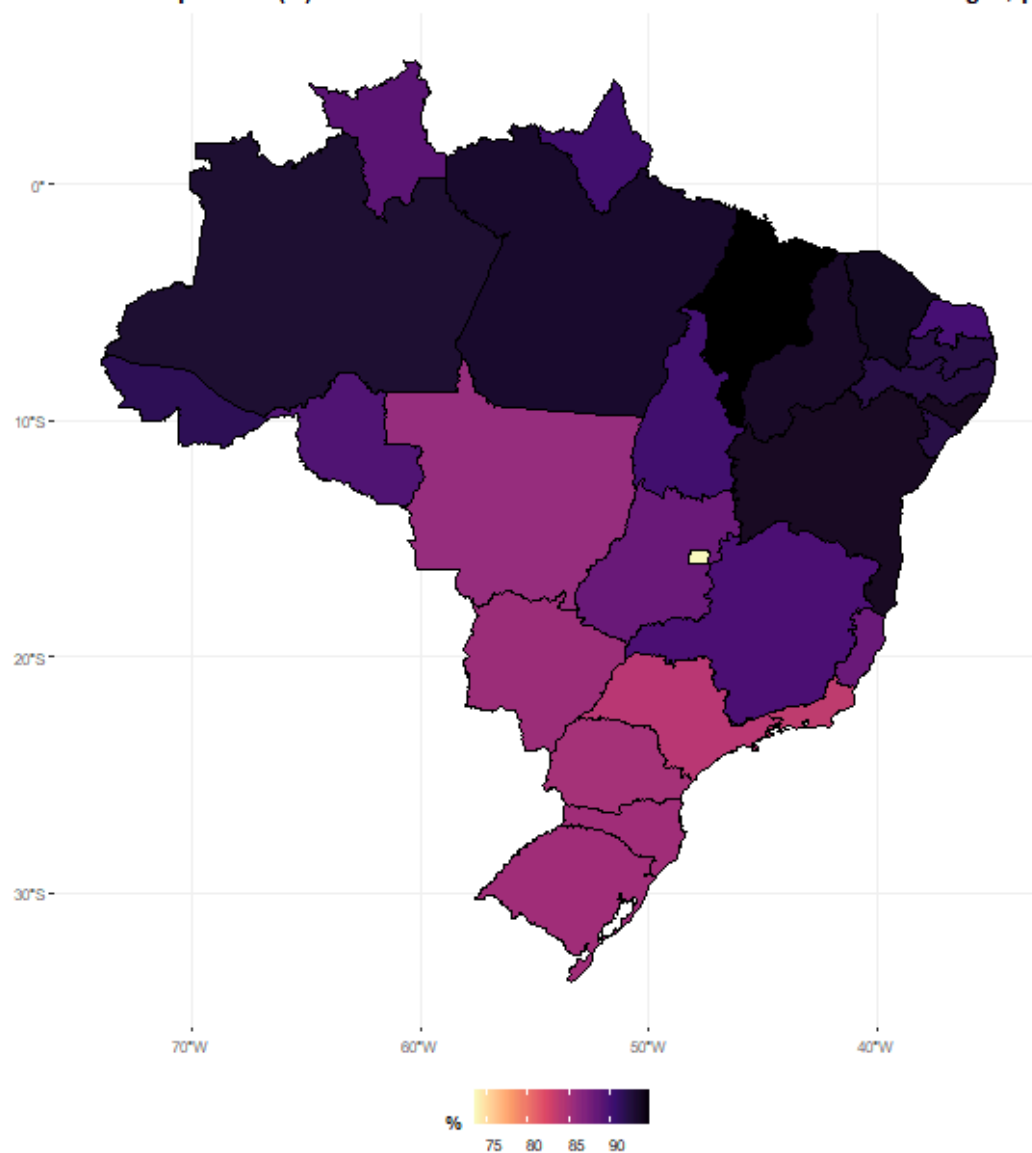
PAUSA!



SOBRE O CASO



Potencial parcela (%) de isentos do IRPFM em 2024 se a lei estivesse em vigor, por UF.



Fonte: IBGE/PNADC, acumulado da primeira visita.



TÓPICO COMPLEMENTAR

Análise a nível do domicílio!



ASPECTOS CONCEITUAIS IMPORTANTES

Estamos falando sobre domicílios, não sobre famílias!



ASPECTOS CONCEITUAIS IMPORTANTES

Estamos falando sobre domicílios, não sobre famílias!

Mais do que isso:

- Estamos tratando de Domicílios Particulares e Permanentes Ocupados (DPPOs)



ASPECTOS CONCEITUAIS IMPORTANTES

Estamos falando sobre domicílios, não sobre famílias!

Mais do que isso:

- Estamos tratando de Domicílios Particulares e Permanentes Ocupados (DPPOs)

O que fica de fora dessa definição?

- Domicílios de temporada
- Domicílios improvisados
- *População não domiciliada*



FORMAS DE ANÁLISES DOMICILIARES USANDO A PNADC

Identificação do domicílio em determinada visita ou trimestre

$$UPA + V1008 + V1014$$

Em que:

- UPA: Unidade Primária de Amostragem
- V1008: Número de seleção do domicílio
- V1014: Número do painel



FORMAS DE ANÁLISES DOMICILIARES USANDO A PNADC

Peso a se utilizar para o domicílio

- Sabemos (Aula 2) que o peso é o mesmo para todos os indivíduos de um determinado domicílio.
- Porém, as características dos indivíduos não serão as mesmas.
- Portanto, a depender da análise a ser realizada, pode impactar nos resultados obtidos.



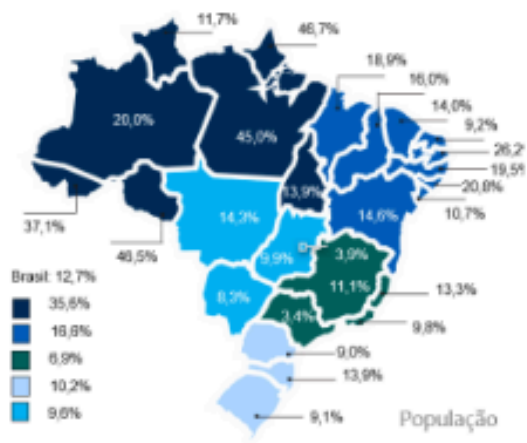
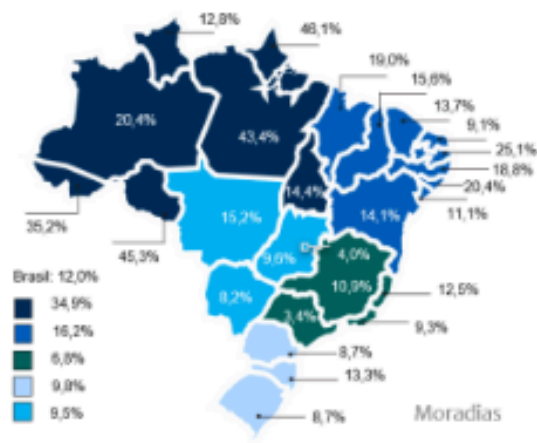
EXEMPLOS DE ANÁLISE A NÍVEL DOMICILIAR



Moradias e população com privação de abastecimento de água por rede geral, em (%) dos totais, 2022

Moradias

População



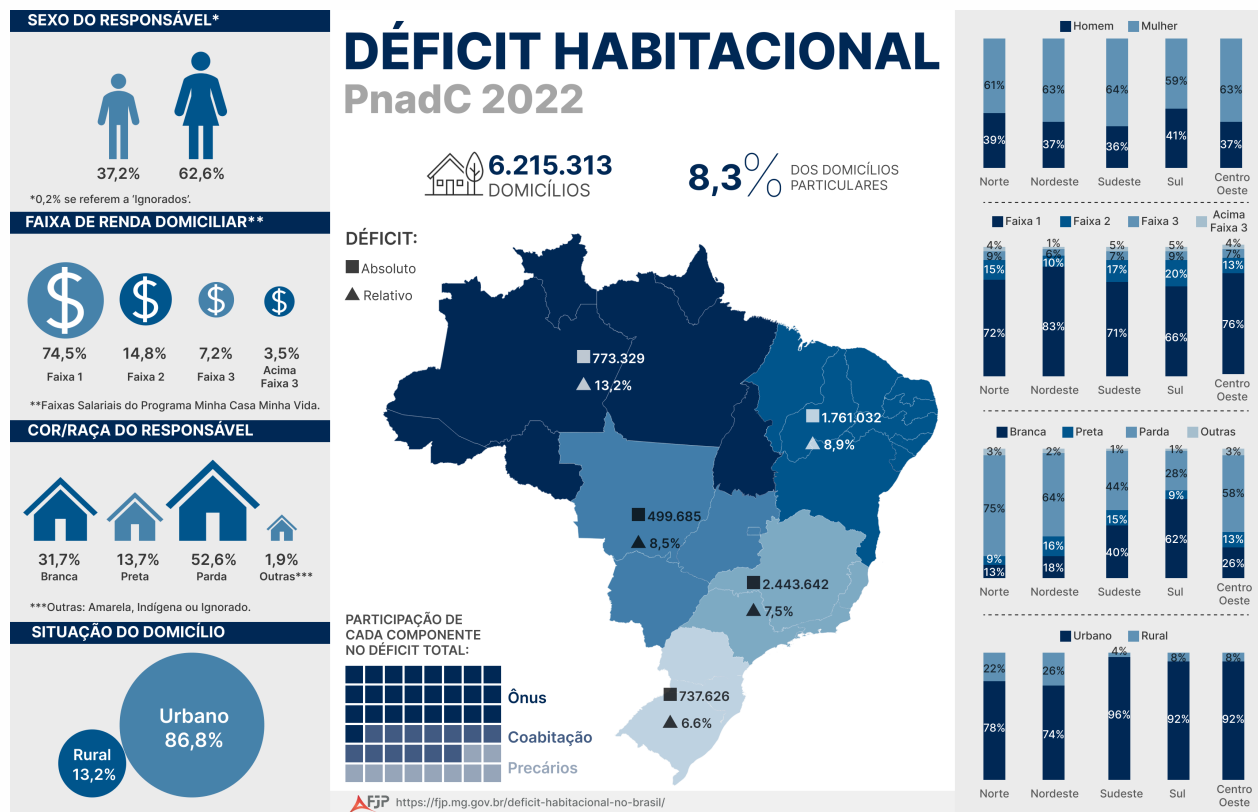
Fonte: PNADC. Elaboração: Ex Ante Consultoria Econômica.

Exemplo 1: Utilizando o bloco de variáveis de habitação (acumulado visita 1)

- Todos os indivíduos do domicílio recebem a mesma resposta.
- Independente do peso que se utilize, estimativas não serão afetadas.



EXEMPLOS DE ANÁLISE A NÍVEL DOMICILIAR



O PESO IMPORTA

Quais estratégias analíticas podemos recorrer?

- **Uso do responsável do domicílio:** usualmente feita¹;
- **Uso de valor médio do domicílio:** média da renda? idade média?
- **Economia de escala intradomiciliar:** diferentes pessoas do domicílio “pesariam” de forma distinta para o valor do domicílio.



PRÓXIMOS PASSOS



RECAPITULANDO

O que vimos no curso:

- Finalidades da PNADC
- Mudanças entre PNAD e PNADC e as vantagens da nova pesquisa
- SIPD e construção da Amostra Mestra
- Operacionalização da coleta da PNADC
- Estratégias de divulgação da pesquisa
- R e RStudio para análise de dados
- Técnicas de avaliação da qualidade das estimativas
- Formas de importação de dados da PNADC em R
- Manipulações e análises da PNADC orientada a problemas empíricos
- Tópico complementar: *análises domiciliares*



NA PNADC

- Limitações de análises
 - Pandemia de Covid-19 e qualidade das estimativas
- Diferentes estimativas para diferentes pesos adotados
- Compatibilização com outras bases de dados públicas ou privadas
- Análises em nível micro (geograficamente)
- Compatibilização da série PNAD (antiga) e PNADC
- Longitudinalização da PNADC
- O que o Censo 2022 traz de novidade?
- Tópicos complementares de análise
 - Classificações de atividade econômica e ocupação (CNAE e CBO)
 - Ponderação manual dos dados da PNADC após manipulação (**exemplo**)
 - Trabalhando com deflacionamentos na PNADC
 - Trabalhando com análises a nível domiciliar com PNADC
- ...



NO R

- Gramática Tydiverse para manipulação de dados
- Gráficos em `{ggplot}`
- Criando suas próprias funções em R
- Mapas em R
- Automatização e reprodutibilidade de processos
- ...



OUTROS CURSOS DO CEBRAP.LAB

- Indicadores sociais e territoriais: como construir e utilizar? - Em novembro
- Análise quantitativa de texto com R - Em novembro



ISSO É TUDO PARA HOJE!



ISSO É TUDO PARA HOJE!

Exercício final!

- Arquivo: [dia 5 > Exercício final.pdf](#)
- Relatório *simples* explorando como um determinado fenômeno de interesse pode ser estudado com a PNADC.
 - Foque aqui em descrever qual tipo de dado da PNADC seria adequado e o motivo, potencialidades e limitações.
- Prazo de envio: 25/10.





CEBRAP

Presidência Adrian Gurza Lavalle

Diretoria Administrativa Victor Callil

Diretoria Científica Arilson Favareto

Coordenação de Seminários Bianca Tavorari

Coordenação de Cursos Monise Fernandes Picanço

Curso

Pesquisa Nacional por Amostra de Domicílios Contínua (PNADC)

Ministrante

Thiago Cordeiro Almeida

E-mail: thiagocordalmeida@gmail.com

Github: [@thiagocalm](#)

