



Explorando os Microdados das Pesquisas Domiciliares Amostrais do IBGE

Gabriel Henrique Oliveira Assunção (IBGE/DPE/COREN/GEMC)

1. O Sistema Integrado de Pesquisas Domiciliares (SIPD)

1. O Sistema Integrado de Pesquisas Domiciliares (SIPD)

- Demanda crescente por informações atualizadas em curto prazo.
- Integrar diversas pesquisas sem sobrecarregar uma única pesquisa com vários temas.
- Consolidação dos conceitos investigados entre as diferentes pesquisas, partilhando seus custos operacionais.
- Padronização de métodos de seleção, esquemas de rotação e estimação.
- Definição do Cadastro Mestre e da Amostra Mestra.

1. O Sistema Integrado de Pesquisas Domiciliares (SIPD)

- Cadastro Mestre é a base com todos os setores censitários elegíveis para composição da Amostra Mestra.
- Setor censitário é a unidade territorial estabelecida para fins de controle cadastral, formado por área contínua, situada em um único quadro urbano ou rural, com dimensão e número de domicílios que permitam o levantamento por um recenseador.
- O Cadastro Mestre abrange todos os setores censitários do território nacional excluindo algumas áreas como agrupamentos quilombolas e indígenas.

1. O Sistema Integrado de Pesquisas Domiciliares (SIPD)

- Amostra Mestra é um conjunto de unidades de área selecionadas de um cadastro, segundo um método probabilístico de seleção.
- A seleção das subamostras das diversas pesquisas pode ocorrer a partir de unidades de área ou de domicílios em todas as unidades de área selecionadas para a Amostra Mestra.
- As subamostras podem ser selecionadas de forma independente ou com controle para obtenção ou não de sobreposição das amostras.
- Domicílios que fizerem parte de uma pesquisa terão um período de descanso antes de serem incluídos em outra.

1. O Sistema Integrado de Pesquisas Domiciliares (SIPD)

- **PNAD Contínua:** investiga as características econômicas e sociais da população, principalmente aquelas relacionadas ao mercado de trabalho.
- **POF:** obtém informações sobre os padrões de consumo e gastos das famílias, servindo de base para avaliação de insegurança alimentar, qualidade de vida, além de ser base para a composição dos índices de preços divulgados pelo IBGE.
- **PNS:** objetiva caracterizar a situação de saúde e os estilos de vida da população (como acesso e uso dos serviços, ações preventivas, continuidade dos cuidados e financiamento da assistência).

2. Planejamento da Amostra Mestra

2. Planejamento da Amostra Mestra

- Unidades Primárias de Amostragem (UPAs) são setores censitários ou conjunto de setores censitários com pelo menos 60 domicílios particulares permanentes (ocupados, fechados, vagos ou uso ocasional).
- Para cada uma dessas unidades são associadas informações sobre divisões administrativas, contagens populacionais e outras características sociodemográficas.
- A principal fonte de informação para o cadastro é o censo demográfico, entretanto, no decorrer de cada ano são realizadas atualizações a partir da Base Operacional Geográfica (BOG).

2. Planejamento da Amostra Mestra

As UPAs foram estratificadas segundo características:

- Administrativas;
- Geográficas e Espaciais;
- De situação (Urbana e Rural);
- Estatísticas.

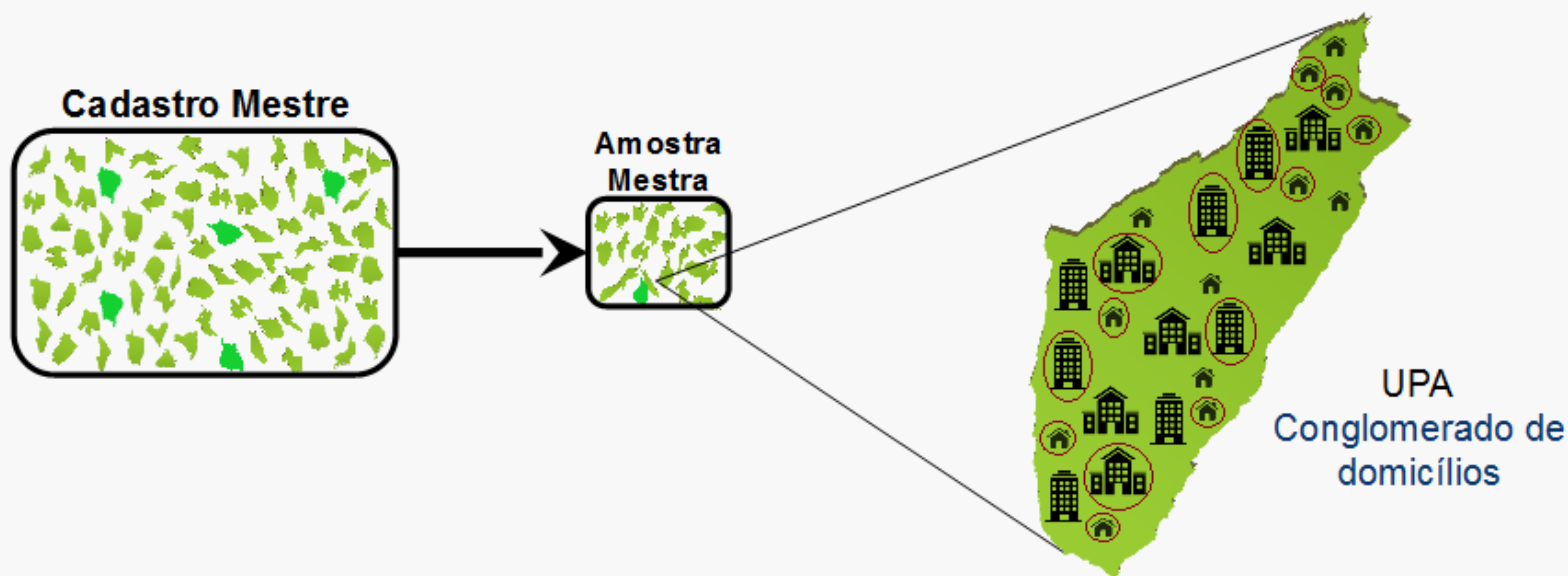
Cada estrato deve ter pelo menos 150 UPAs.

2. Planejamento da Amostra Mestra

- A estratificação das UPAs leva em consideração os objetivos de todas as pesquisas do SIPD.
- Permite o controle da seleção e do tamanho da amostra para possíveis domínios de divulgação.
- Garante espalhamento da amostra no território.
- Controla a precisão das estimativas por domínio de estimação.
- Busca facilitar as operações de controle de coleta, além de reduzir custos.

2. Planejamento da Amostra Mestra

- O plano amostral é conglomerado em dois estágios com estratificação das UPAs.
- No primeiro estágio ocorre a seleção de UPAs com probabilidade proporcional ao número de domicílios.
- Já no segundo estágio ocorre a seleção dos domicílios por amostragem aleatória simples que farão parte da pesquisa.



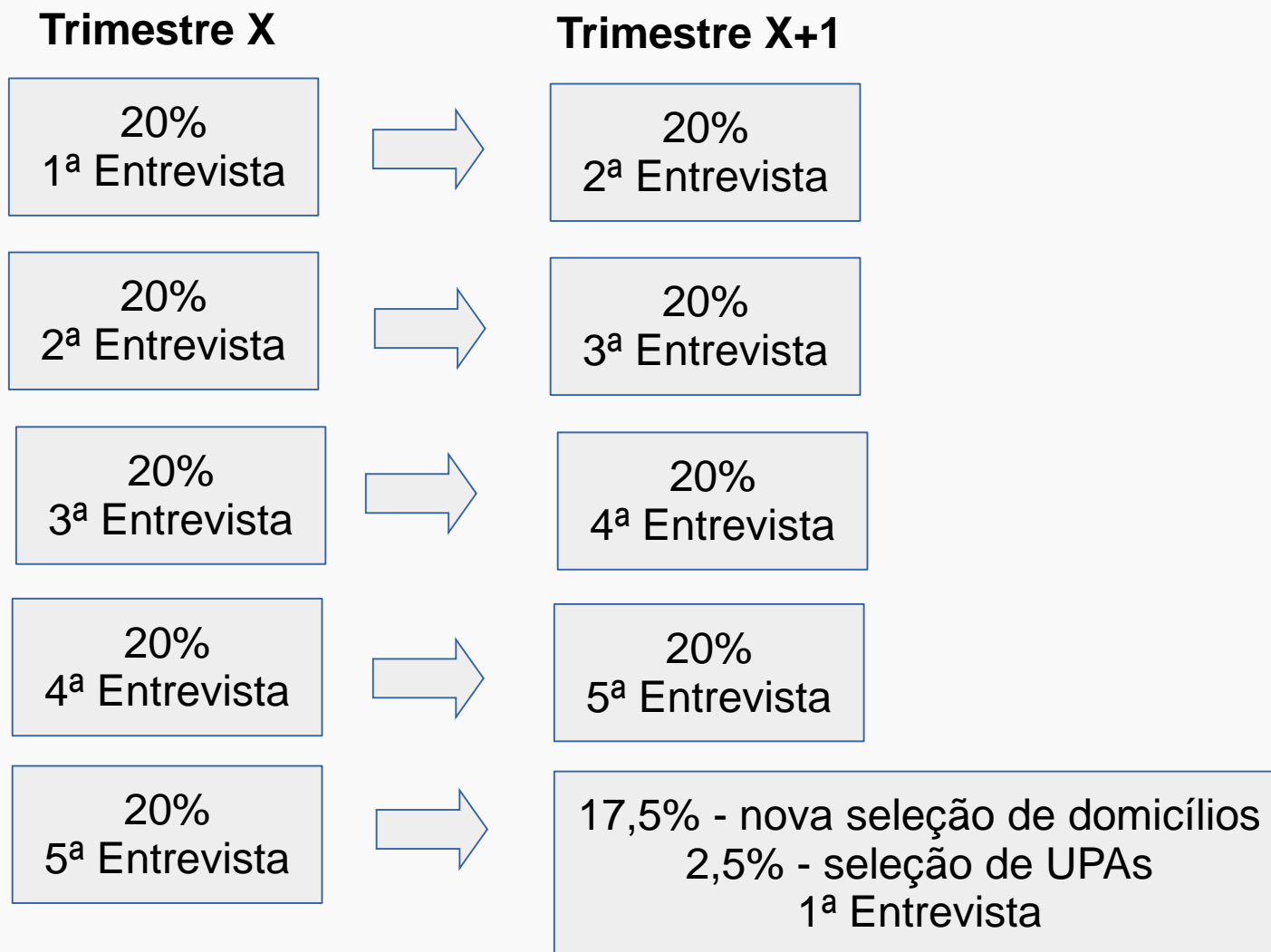
2. Planejamento da Amostra Mestra

Como a PNAD Contínua é a principal pesquisa do SIPD para definir o tamanho de amostra é preciso levar em consideração alguns aspectos de sua formulação:

- Pesquisa trimestral;
- Rotação da amostra de UPAs, com sobreposição de domicílios em trimestres consecutivos;
- Esquema de rotação escolhido: 1 mês - 2 meses (5 entrevistas);
- Criação de 15 grupos para operacionalizar o esquema de rotação, cada um possuindo 8 rotações;
- Seleção de 14 domicílios em todas as 15.096 UPAs da Amostra Mestra, totalizando 211.344 domicílios por trimestre.

2. Planejamento da Amostra Mestra

PNAD Contínua



2. Planejamento da Amostra Mestra

- Como a PNAD Contínua é uma pesquisa por amostragem probabilística, para obtenção dos resultados de maneira correta é imprescindível a associação de um fator de expansão ou peso a cada unidade selecionada (UPAs, domicílios e moradores).
- Para os pesos de cada unidade são levados em consideração as probabilidades de seleção, os ajustes por não resposta e de calibração.
- Há ainda, peso trimestral, utilizado para as estimativas produzidas trimestralmente, e peso anual, quando os indicadores têm por referência o ano.
- Para os pesos das UPAs produzidos trimestralmente, são consideradas a probabilidade de seleção da UPA no grupo de rotação e a razão dos domicílios particulares permanentes ocupados e fechados da UPA em relação as demais UPAs do grupo de rotação.

2. Planejamento da Amostra Mestra

- Para as estimativas calculadas apenas uma vez por ano, com acumulação das amostras trimestrais não sobrepostas de UPAs, o cálculo dos pesos leva em conta a utilização de apenas 80% da amostra.

Mês	J	K	L	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	A	B	C	D	E	F	G	H	I
10	5			4			3			2			1														
11		5			4			3			2			1													
12			5			4			3			2			1												
1				5			4			3			2			1											
2					5			4			3			2			1										
3						5			4			3			2			1									
4							5			4			3			2			1								
5								5			4			3			2			1							
6									5			4			3			2			1						
7										5			4			3			2			1					
8											5			4			3			2			1				
9												5			4			3			2			1			
10													5			4			3			2			1		
11														5			4			3			2			1	
12															5			4			3			2			1

2. Planejamento da Amostra Mestra

- Para a definição do peso por domicílio, primeiramente é calculado o peso básico que é dado pelo inverso da probabilidade de seleção do domicílio, que é igual em cada UPA, e pode ser decomposto na parcela correspondente à seleção da UPA e na parcela correspondente à seleção dos domicílios dentro da UPA.
- Após a determinação do peso básico, ele é ajustado para compensar as perdas de entrevistas por não resposta, ou seja, entrevistas não realizadas por recusa do informante, por não contato com o morador ou por outro motivo para perda em domicílios ocupados.
- Por fim, com o intuito de melhorar a qualidade das estimativas é realizada a calibração, onde os pesos são ajustados para que ao estimar o total populacional de certos níveis geográficos, a estimativa obtida coincida com a estimativa populacional do dia 15 do mês do meio do trimestre de referência nas estimativas trimestrais e no dia 01 de Julho para estimativas anuais.

2. Planejamento da Amostra Mestra

- Os principais indicadores da pesquisa são totais de características de interesse ou razões entre totais de duas características, como taxas. Outros indicadores de interesse em pesquisas como a PNAD Contínua dizem respeito a diferenças dos indicadores em dois períodos de tempo subsequentes.
- As estimativas de variância são obtidas usando o método do Ultimate Cluster, também conhecido como método do Conglomerado Primário, sendo que para estimar a variância dos estimadores de razão é utilizada uma aproximação da variância de uma razão através de Linearização de Taylor.
- Quando o estimador é uma diferença entre indicadores trimestrais, a sua variância é composta pelas variâncias dos dois indicadores trimestrais e pela covariância entre estes indicadores. A sobreposição da amostra em trimestres subsequentes melhora a precisão da estimativa das diferenças, pois quanto maior for a sobreposição, espera-se que a covariância seja também maior, reduzindo a variância da estimativa.

3. Pacotes para Leitura dos Microdados

3. Pacotes para Leitura dos Microdados

- As pesquisas domiciliares amostrais do IBGE envolvem metodologias de amostragem que levam a necessidade de incorporação do desenho amostral para obtenção das estimativas desejadas.
- Para incorporação do desenho amostral devem ser considerados os pontos abordados durante esta apresentação.
- Criação de pacotes para facilitar a leitura dos microdados das pesquisas domiciliares amostrais do IBGE.
- Pacotes disponibilizados no software R através do repositório CRAN.

3. Pacotes para Leitura dos Microdados

Por que utilizar o software R?

- É um software gratuito.
 - Possui código fonte aberto.
 - Grande variedade de métodos e pacotes.
 - Atualizações simples e rápidas.
 - Capaz de considerar todo o plano amostral das pesquisa para realizar as análises corretamente.
- Existe ainda a possibilidade de exportação dos microdados para um formato compatível com outro software (SAS, Stata ou SPSS) através do pacote *haven*.

3. Pacotes para Leitura dos Microdados

Pacotes elaborados pelo IBGE para leitura dos microdados das pesquisas domiciliares amostrais no R:

- **PNADcIBGE:** permite baixar e preparar os microdados da PNAD Contínua para análise.
- **COVIDIBGE:** permite baixar e preparar os microdados da PNAD COVID19 para análise.
- **PNSIBGE:** permite baixar e preparar os microdados da PNS para análise.
- **POFIBGE:** permite baixar e preparar os microdados da POF para análise. [Em fase de desenvolvimento]
- **SIPDIBGE:** pacote de coletânea para controle e manutenção dos pacotes das pesquisas contempladas pelo SIPD.

3. Pacotes para Leitura dos Microdados

Pacotes que possibilitam realizar a análise dos microdados das pesquisas domiciliares amostrais no R:

- **survey**: pacote que permite análise e modelagem de dados provenientes de pesquisas com amostras complexas.
- **convey**: pacote para análise de concentração de renda para amostras complexas.
- **lavaan.survey**: pacote para modelos de equações estruturais para amostras complexas.
- **surf**: pacote para cálculo de fluxos brutos utilizando amostras repetidas em pesquisas com amostras complexas.

3. Pacotes para Leitura dos Microdados

- Os pacotes foram desenvolvidos com o intuito de facilitar a importação, leitura e análise dos microdados das pesquisas.
- Apresentam funções bem simples e métodos que possibilitam o uso destas informações de maneira correta.
- Para a PNAD Contínua existe a possibilidade de realizar a leitura dos microdados trimestrais e anuais (acumulados em determinada visita ou concentrados em determinado trimestre).
- Para a PNS e a POF o pacote permite a leitura dos microdados completos, para os módulos de morador selecionado e também para módulo específicos, quando existir, como Antropometria.
- Para a PNAD COVID19 o pacote permite a leitura somente dos microdados mensais.

3. Pacotes para Leitura dos Microdados

- Os pacotes permitem a leitura dos microdados de forma online/automática direto do site do IBGE e também de forma offline/manual de um determinado diretório do computador.
- Existe a possibilidade através de uma função do pacote para rotulação das variáveis categóricas de acordo com os dicionários das pesquisas.
- Está disponível também uma função para disponibilização dos índices para deflacionamento das variáveis de rendimento, incluindo um manual para auxílio aos usuários.
- Por fim, existe uma função própria para incorporação do desenho amostral das pesquisas que é parte fundamental para obtenção correta das estimativas.

Agradecemos a todos pela atenção!

Estamos a disposição para quaisquer dúvidas!



Contatos:



Gabriel Henrique Oliveira Assunção

pacotesipd@ibge.gov.br

gabriel.assuncao@ibge.gov.br